REPORT RESUMES

ED 002 958   24
THE NATURE OF ANALYSIS AND SYNTHESIS AND SOME CONDITIONS IN THE CLASSROOM WHICH FACILITATE OR RETARD THESE COGNITIVE PROCESSES--PART ONE, SOME CONDITIONS WHICH AFFECT COGNITIVE BEHAVIOR.
BY- ROKEACH, MILTON AND OTHERS
MICHIGAN ST. UNIV., EAST LANSING
REPORT NUMBER CRP-879-PART-1   PUB DATE   65
CONTRACT OEC-SAE-8830
EDRS PRICE MF-$0.36 HC-$9.12   228P.

DESCRIPTORS- *COGNITIVE ABILITY, COGNITIVE MEASUREMENT, CLASSROOM RESEARCH, COMPARATIVE ANALYSIS, COLLEGE INSTRUCTION, *GROUP INSTRUCTION, *INDIVIDUAL INSTRUCTION, *PROGRAMED INSTRUCTION, *CONVENTIONAL INSTRUCTION, LEARNING ACTIVITIES, PROBLEM SOLVING, TASK PERFORMANCE, DENNY DOODLEBUG PROBLEM, EAST LANSING, MICHIGAN

COLLEGE CLASSROOM CONDITIONS WHICH ADVANCE OR IMPEDE STUDENT ABILITY IN COGNITIVE ANALYSIS AND/OR SYNTHESIS BY THOUGHT AND PERCEPTION WERE EXPLORED. TWO BASIC COMPARISONS WERE MADE--(1) INDIVIDUAL VERSUS GROUP MEMORIZATION AND PROBLEM SOLVING USING POOLING-OF-ABILITIES MODELS AND (2) THE EFFECTIVENESS OF PROGRAMS VERSUS CONVENTIONAL INSTRUCTIONAL METHODS ON LEARNING AND PROBLEM SOLVING. IN THE FIRST INSTANCE THE MODEL TESTED HYPOTHESIZED THAT THE PROBABILITY OF A GROUP SOLVING A PROBLEM WAS GREATER THAN FOR AN INDIVIDUAL AND THAT GROUP SUPERIORITY IS NOT NECESSARILY THE RESULT OF ANY FACILITATIVE QUALITIES OF THE GROUP. STUDENT GROUPS OF TWO AND THREE PARTICIPANTS, AS WELL AS INDIVIDUALS WERE TESTED USING AS THE TASKS--(1) THE RECALL OF NONSENSE WORDS, AND (2) THE JOE DOODLEBUG PROBLEM (THE RECALL OF THE FUNCTIONING AND MOVEMENT OF AN IMAGINARY BUG). GROUPS OF TWO PERFORMED BETTER ON THE FIRST PROBLEM, WHILE GROUPS OF THREE EXCELLED ON THE SECOND. IN BOTH EXPERIMENTS THERE WAS NO INDICATION OF GROUP FACILITATION, AND SOME INTERFERENCE WAS APPARENT IN ALL GROUP PERFORMANCES. THESE DATA, THEREFORE, COULD NOT BE USED TO REJECT THE HYPOTHESIS OF THE EXPERIMENTAL MATHEMATICAL MODEL. THE DOODLEBUG PROBLEM WAS USED AGAIN TO TEST THE EFFECTIVENESS OF PROGRAMED AND TRADITIONAL INSTRUCTION. THE SUBJECTS PARTICIPATING IN THE TESTS WERE INTRODUCTORY PSYCHOLOGY STUDENTS IN COLLEGE AND EIGHTH-GRADE STUDENTS FROM TWO JUNIOR HIGH SCHOOLS. IN GENERAL, TRADITIONAL SUBJECTS SHOWED BETTER LEARNING RESULTS AT BOTH GRADE LEVELS, PARTICULARLY FOR STUDENTS WHO WERE GIFTED. (JH)

ED 002958 14

The Nature of Analysis and Synthesis and Some Conditions in the Classroom Which Facilitate or Retard These Cognitive Processes

PART ONE: SOME CONDITIONS WHICH AFFECT COGNITIVE BEHAVIOR

A first report of research submitted in January, 1963, supported
in whole or in part by the United States Office of Education,
Department of Health, Education, and Welfare.

Cooperative Research Branch Project No. 879 (SAE 8830)
July 1, 1960 to June 30, 1965

Principal Investigator: Milton Rokeach, Michigan State University

ED 002958

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

TABLE OF CONTENTS

Page

1. INDIVIDUAL AND GROUP MEMORIZATION AND PROBLEM SOLVING: A TEST OF THE LORGE-SOLOMON POOLING-OF-ABILITIES MODEL 1
by Ronald A. Hoppe

2. MEMORIZING BY INDIVIDUALS AND GROUPS: A TEST OF THE POOLING-OF-ABILITY MODEL 95
by Ronald A. Hoppe

3. THE EFFECTIVENESS OF PROGRAMMED AND TRADITIONAL METHODS OF INSTRUCTION IN A COMPLEX VERBAL PROBLEM SOLVING SITUATION 104
by Walter A. Watman

4. A COMPARISON OF THE EFFECTS OF PROGRAMMED AND CONVENTIONAL METHODS OF INSTRUCTION ON LEARNING AND PROBLEM SOLVING IN OLDER CHILDREN 132
by Marilyn Daugherty

INDIVIDUAL AND GROUP MEMORIZATION AND PROBLEM-SOLVING: A TEST OF THE LORGE-SOLOMON POOLING-OF-ABILITIES MODEL

By

Ronald A. Hoppe

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1962

3414a
(125)

ACKNOWLEDGMENTS

I am very appreciative of the guidance and beneficial assistance of my major professor, Dr. Milton Rokeach.

I offer many thanks to my committee, Drs. M. Ray Denny, Paul Bakan and Stanley Ratner who had many important suggestions.

Also, I am deeply grateful to Dr. Frank Restle whose instruction was the impetus for this work.

R. A. H.

PRECEDING PAGE BLANK-NOT FILMED

TABLE OF CONTENTS

CHAPTER Page

I. INTRODUCTION 1

II. THEORY 18

III. METHOD 28

IV. RESULTS 35

V. CONCLUSIONS 82

VI. SUMMARY 89

BIBLIOGRAPHY 91

LIST OF TABLES

TABLE Page

2.1 Obtained and predicted means of Perlmutter and DeMontmollin 23

4.1 Obtained recall results: A comparison of the overall variances of individuals, groups of 2 and groups of 3 . 37

4.2 Obtained recall results: Comparisons of obtained trial variances 38

4.3 Analysis of variance of obtained recall data 39

4.4 T-test comparison of obtained recall results 41

4.5 Recall, total means--predicted and obtained. 42

4.6 Obtained vs. predicted recall results trial by trial for groups of 2 47

4.7 Obtained vs. predicted recall results trial by trial for groups of 3 48

4,8 First trial proportions for individuals, groups of 2 and groups of 3 and predicted proportions correct for groups of 2 and groups of 3 49

4.9 Obtained Doodlebug results: Proportions of individuals, groups of 2 and groups of 3 passing solution, analysis and synthesis measures. 52

4.10 Obtained Doodlebug results: Individuals vs. groups of 2, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis. 54

4.11 Obtained Doodlebug results: Individuals vs. groups of 3, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis. 60

LIST OF TABLES - Continued

TABLE Page

4.12 Obtained Doodlebug results: Groups of 2 vs. groups of 3, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis. . 62

4.13 Obtained vs. predicted Doodlebug results: Arcsin test of the difference between obtained and predicted proportions of groups of 2 solving the Doodlebug problem and passing measures of analysis and synthesis. 64

4.14 Obtained vs. predicted Doodlebug results: Arcsin test of the difference between obtained and predicted proportions of groups of 3 solving the Doodlebug problem and passing measures of analysis and synthesis. 65

4.15 Obtained vs. predicted Doodlebug results: Kolmogrov-Smirnov one sample test of differences between obtained and predicted distributions of groups of 2 76

4.16 Obtained vs. predicted Doodlebug results: Kolmogrov-Smirnov one sample test of differences between obtained and predicted distributions of groups of 3 78

LIST OF FIGURES

FIGURE Page

4.1 Obtained recall results. Groups of 2, groups of 3 and individuals. 44

4.2 Obtained and predicted recall results of groups of 2. . 45

4.3 Obtained and predicted recall results of groups of 3. . 46

4.4 Obtained Doodlebug results. Cumulative distributions of proportions of solutions in groups of 2, groups of 3 and individual conditions 55

4.5 Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2, groups of 3 and individuals overcoming 1 belief 56

4.6 Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2, groups of 3 and individuals overcoming 2 beliefs 57

4.7 Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2, groups of 3 and individuals solving after overcoming 1 belief 58

4.8 Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2, groups of 3 and individuals solving after overcoming 2 beliefs 59

4.9 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving 66

4.10 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 overcoming 1 belief 67

LIST OF FIGURES - Continued

FIGURE

Page

4.11 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 overcoming 2 beliefs. 68

4.12 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving after overcoming 1 belief 69

4.13 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving after overcoming 2 beliefs 70

4.14 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 solving. 71

4.15 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 overcoming 1 belief. 72

4.16 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 overcoming 2 beliefs 73

4.17 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 solving after overcoming 1 belief. 74

4.18 Obtained vs. predicted Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 solving after overcoming 2 beliefs 75

ABSTRACT

INDIVIDUAL AND GROUP MEMORIZATION AND PROBLEM-SOLVING: A TEST OF THE LORGE-SOLOMON POOLING-OF-ABILITIES MODEL

by Ronald A. Hoppe

A mathematical model was used to examine group performance in two experiments. The model was constructed by Lorge and Solomon and states that the probability of a group solving a problem is greater than the probability of an individual solving a problem simply because there are more individuals in a group. Therefore, group superiority may occur without any facilitative effects of group interaction. Furthermore, the model allows for a prediction of group behavior or a baseline against which to judge whether or not there is facilitation or interference in group interaction.

The first experiment tested the model using as the task the recall of nonsense words. Groups of 2, groups of 3 and individuals were given 9 recall trials on 8 nonsense words. In this experiment to apply the Lorge-Solomon model each nonsense word was treated as a problem to solve. The results of this experiment demonstrated that groups of 2 and groups of 3 were superior to individuals. Groups of 3 were not quite significantly superior to groups of 2. Also, in all conditions there was a strong serial position effect on the first trial. When this serial position effect was taken into account the predictions of groups of 2 and 3 were quite accurate. However, the model slightly over predicted the performance of groups of 3 which may indicate a tendency for

interference in groups of this size. It was concluded that an explanation of group superiority which uses the facilitative effects of groups is not suitable since the superiority can be explained more simply by the Lorge-Solomon model.

In the second experiment groups of 2, groups of 3 and individuals were given a complex problem (Doodlebug problem) to solve. This problem allowed for a separation of the analysis and synthesis phases of problem solving. The results showed groups of 3 to be superior to individuals and groups of 2 in analysis, synthesis and solving of the problem. Groups of 2 were not significantly different from individuals. The Lorge-Solomon model considerably over predicted the performance of groups of 2 in analysis, synthesis and solution to the Doodlebug problem. This indicated the presence of interference in groups of this size. The predictions of the performance of groups of 3 was accurate except for the synthesis phase of problem solving. In this phase groups of 3 performed more poorly than predicted by the Lorge-Solomon model suggesting interference. However, the interference was not strong enough to disrupt the prediction of the solutions by groups of 3.

The strong interference present in groups of 2 seemed to be due to one person, regardless of his problem solving ability, dominating the solution to the problem. This notion was supported by comparing expected and obtained participation values.

Again, in the second experiment there was no indication of group facilitation. Group superiority, where it occurred, could be explained by the Lorge-Solomon model.

In both experiments there was some indication of interference in group performance. In the first experiment the interference was slight. In the second experiment the interference was strongest in groups of 2 and present somewhat in groups of 3. No evidence was found for group facilitation.

INTRODUCTION

The comparison of group and individual performance has a long and varied history. Studies have compared group and individual performance on a variety of simple and complex tasks. The present study also compares group and individual performance, and, furthermore, attempts to predict group performance from independent individual performance.

One general result when comparing groups and individuals on problem solving ability is that the group often is superior to individuals (Kelly and Thibaut, 1954). In 1955, Lorge and Solomon devised a mathematical model to account for group superiority. The model says that the probability of a group solving a problem is greater than the probability of an individual solving the same problem simply because there is a greater number of individuals working on the problem in a group. When more than one person is working on a problem the probability of any one of the persons solving the problem is greater than the probability of an individual working independently solving the problem. This model assumes that individuals will work the same in a group as they do individually and the group neither facilitates nor inhibits the individual's performance. The group is superior to individuals just because there are more individuals in the group. The model provides a baseline for us to judge what is occurring in groups. If there is no facilitation or inhibition, groups should perform as the model predicts. A deviation from the prediction may indicate either facilitation or inhibition depending on whether groups are superior or inferior to the prediction. Therefore, the model can be used as a tool to study what is occurring in groups.

The present thesis involves two experiments contrasting groups and individuals. Both experiments will use the Lorge and Solomon model to predict the group behavior. The first experiment concerns giving groups and individuals a series of trials on a series of nonsense words and it will be their task to recall as many words as possible. From the individual results the group results will be predicted. A modification of the Lorge and Solomon model is necessary so that it will apply to recall instead of problem solving. This is discussed in the next chapter. The second experiment uses a more complex task--the Doodlebug problem (Rokeach, 1960)--and, again the Lorge and Solomon model will be used to predict group performance from individual performance. The Doodlebug problem is unique because it permits a study of two separate phases, analysis and synthesis, of the problem solving process.

Our purpose is to try to understand by what process individual products become group products. Is there a simple pooling which the Lorge and Solomon model suggests, or is there facilitation or inhibition in groups? Furthermore, perhaps pooling is an adequate explanation for simple tasks such as recall, but is inadequate for more complex tasks such as a difficult, unique problem--or vice versa.

Before delving into the specific mathematical and psychological theory involved in these two experiments, a review of relevant past research is in order. One type of early research was to compare individual results with grouped individual results. In this case the group was simply the grouping of individual data and was not a face to face interacting group. Gordon's research (1924) is an example of this. She correlated individuals' rankings of weights with the actual ranking of the weights and found a mean correlation of .41. Next, she combined individual rankings in various size groups (ranging from 5 to 50) and found the mean rankings for all the individuals in each of the groups. These she correlated with the actual rankings and found that the average correlations went as high as .94 for 50 member groups. She gave this as evidence for the superiority of group judgments. Stroop (1932), however, demonstrated that Gordon was simply verifying the Spearman-Brown prophecy formula. Stroop let one individual rank the weights for the same number of times as Gordon had individuals in groups. He obtained essentially the same results as did Gordon and concluded that the superiority she found in groups is a statistical artifact. Johnson (1955) and Lorge, Fox, Davitz, and Brenner (1958) made a similar conclusion. While these two early studies do not involve face to face interacting groups, they do serve to point out the necessity of controlling for just the additional numbers found in groups.

Watson (1928) gave individuals and groups a number of words, such as "educators" and "neurotics" from which it was their task to make as many other words as possible. His results indicated that groups produce more words than individuals and also more than the best individuals. However, he found that the summed words for the same number of individuals working separately as had worked together in groups were more than those produced by the groups. Anderson (1961) repeated Watson's experiment because he felt that Watson did not give his groups enough time to allow the groups' superiority over the sum of individuals to evidence itself. Anderson gave his groups more time and found that the groups were not inferior to the sum of individuals. Instead, he found no significant difference between groups and the sum of individuals.

Watson and Anderson have controlled, experimentally, for what has become known as "pooling" by summing the results of the same number of subjects who have worked separately on the problem. Watson chose not to let this enter into his explanation of what may have been occurring in the group interaction and Anderson found that his real groups were not superior to the pooled individuals but did not suggest why they did not differ.

The notion of pooling suggests that the group product is a result of the pooling of the accomplishments of the individual group members. This is in accordance with Allport's (1962) suggestion that the group product is a product of individuals, nothing more. In the Watson (1928) and Anderson (1961) experiments pooling may be considered as the summation of individual products, whereas, in problem solving (to be discussed shortly) it can be considered as the best individual product resulting in the group product.

Usually attached to the notion of pooling is the assumption that the individuals working in a group setting are neither better nor worse than they are when working alone, presuming that the group interaction neither facilitates nor interferes with individual performance. The proper experimental construction of pooling has been done using artificial groups which have been made from subjects who work separately and independently and whose results are then combined. The same number of individual results as there are subjects working together in a real group are combined. This combination of individual results has been labelled a "nominal group" (Faust, 1959). Nominal groups can be used to test the pooling notion or as a control, for if facilitative effects are occurring in a group, then the real groups should be superior to nominal groups. Conversely, if there is interference, the real groups should perform more poorly than nominal groups.

The notion of pooling has had a confused history. An attempt to clarify the notion will be made a bit later. Many studies have neglected the pooling control when comparing group and individual products and some of the experimenters who made use of the control did not interpret their results in terms of pooling.

Unfortunately, the emphasis on group superiority has led the researchers to explain the superiority without worrying about by what process the group product is derived from the individuals in the group.

In Husband's (1940) study the emphasis on group superiority has been somewhat more fruitful. He gave pairs of subjects and single subjects a variety of problems and found pairs to be superior to individuals on some of the problems. Through observations of the pairs he concludes that the superiority is due to a division of labor. Often, each member of a pair would be working independently on part of the task. This suggests that the group product might result from a pooling of the individual products; however, this was not studied experimentally.

Gurnee (1937a) offers a few notions relevant to the explanation of group products. In one study individuals and groups judged the truth and falsity of numerous statements. In this study the subjects judged each statement first as individuals and then as a group. The group judgment, however, was not a product involving interaction of the members since the individuals expressed their judgments either by acclamation or by raising their hands, rather than discussing the statement and emerging with one agreed judgment. He found the groups superior to the average individual judgments and also to a pooling of the individual judgments which were made just prior to the group judgments. Gurnee suggests that the superiority might be due to the subjects making two responses and the second response being more accurate than the first, but he also feels that social stimuli had an influence and the doubtful subjects would hold back their response just long enough to observe the dominant side and then vote accordingly. However, Zajonc (1962a) using Gurnee's (1937a) individual results and a prophecy formula similar to the Spearman-Brown was able to predict his group results. This indicates that the social stimuli had no effect.

Thorndike (1938) performed a somewhat similar experiment but included social interaction in the group process. He found that individual subject's judgments improved after group discussion, but, again, just how much the subject's judgments would change due to making more than one judgment is unknown since the proper controls were not applied in Thorndike's experiment.

Other studies which have involved a variety of tasks have shown a change in individual judgment after or during group processes. Asch (1952) has demonstrated a change in the judgment of the length of lines; Sherif (1935) has shown a change using the autokinetic effect and group judgment; Lewin (1947) has shown a change in food choices; Back (1951) has shown a change in interpretation of pictures, and Coch and French (1948) have shown a change in job satisfaction. These studies are generally relevant to the understanding of group processes but are not particularly relevant to understanding the group product in problem solving and recall other than to point out that individuals may change their behavior as a result of group processes.

Jenness (1932) had 4 groups of subjects make judgments of the number of beans in a bottle. Three of the groups participated in group discussion between the two judgments. The fourth group, a control, merely made two judgments without discussion. In one group the judgments of the individuals improved more than the control group's judgments. In this group individuals were informed as to what the member's first judgment was. The mean was used to compare the groups, so Jenness' control group does not amount to a control for pooling, but is a control for an "averaging" influence.

In a second study by Gurnee (1937b) individuals and groups learned a stylus maze. The group members were required to vote at each turn of the maze. Gurnee found groups to be superior to individuals. In this study he emphasizes the "scattering of individual errors" to account for the superiority of groups which may imply a kind of pooling in the group. However, in a further study (Gurnee, 1939) which was very similar to the one just mentioned, Gurnee suggests that the social rewards in the group situation are very important. In this study Gurnee demonstrated that the individual members of groups learned more in the same number of trials than did individuals who worked alone on the task. The social rewards which Gurnee refers to are the comments both oral and silent made by the individuals in the groups to a vote made by another member. Subjects might groan, nod, etc. to a vote by a member to turn the stylus in a certain direction. These rewards, both positive and negative, increase the learning in the group setting according to Gurnee.

In these last two studies Gurnee obtained introspective data about the subject's motivational level and he reports no difference between those individuals working in groups and those working as individuals.

From Gurnee's studies there are two suggestions of how the group product is derived from the individual products. First, there is the notion of pooling which Gurnee discussed in rather vague terms, and secondly, there is the notion of social rewards affecting the individual members so that they may behave differently in group situations than they do when performing alone which would cause the individual products available for pooling to be different.

Shaw (1932) focuses on explaining group superiority. She gave "eureka" problems to groups of 4 and also to individuals. They are labelled "eureka" problems because to each of them there is a unique answer which seems to be one of sudden insight. A greater proportion of the groups solved the problems within the given time limit than did the individuals, and Shaw explained this superiority by saying that the groups were more efficient in rejecting wrong answers than were the individuals, but there was no explanation of how the group product results from the individuals in the group.

Timmons' (1942) somewhat later study involves a few notions of how individual effects might result in group products. First, he considered an averaging influence, that is, the group behaves similarly to the average of the individuals in the group. Secondly, he considered the majority influence, that is, the group behaves as does the majority of its members. In his experiment he gave individuals and groups of 4 a series of parole problems. To each of the problems the group and individuals were required to rank from best to worst a number of solutions to the problems. The groups discussed the problems and the individuals read a pamphlet about parole problems. Timmons studied the averaging effect by averaging the results of 4 persons working as individuals and comparing their results with the results of the real groups. He found that the real groups were superior to the averaging influence. He also examined the majority influence by looking at the answers of the 4 individuals who worked separately and by taking the ranking which was selected most often by the individuals. He compared results using the majority rankings with the groups' rankings and again found the group superior. Finally he combined both the averaging and majority influences and found that the group was superior to both. Timmons concluded that while both these influences may have been operating in the groups, they do not explain all of the group superiority and that "factors inherent" in the group discussion account for "much" of the superiority of the group.

A more recent experiment by Barnlund (1959) is also directed at studying the influence of the majority. He had two comparable lists of multiple choice problems, one of which the subjects worked on as individuals and the other in groups of from 3 to 6 members. His results showed the groups superior to individuals in the number of problems solved correctly. He found that the majority influence did not account for the group superiority. In addition, he tested the superior man notion. In his groups he found one person who was superior to the others on the basis of the individual scores. Using this superior person's results on the list which the members had worked as individuals, he still could not account for the superiority of the groups. He concluded that group decisions reached through "co-operative deliberation" were superior to decisions made by individual members working alone and to majority rule. He said that this superiority was due to "psychological factors inherent in discussion"; a conclusion very similar to the one reached by Timmons, and one wonders what the inherent factors could be.

Shaw, (1932), Jenness (1932), Timmons (1942) and Barnlund (1959) in their studies did not use a simple pooling control. Timmons used control groups of the same number of individuals working on the problem as he had individuals in interacting groups but he looked at averages and the majority influence rather than pooling.

Perlmutter and DeMontmollin (1952) and Perlmutter (1953) have studied group and individual recall. In the first study groups and individuals were given 5 trials on a list of 19 nonsense words. The groups discussed the possible words at the end of each presentation. Perlmutter and DeMontmollin found that the groups were superior throughout the 5 recall trials and concluded that this superiority resulted from interaction among the members. This interaction, they submitted, facilitated the individual member's recall. They also noted that the groups exercised a critical and evaluating function similar to that which Shaw mentioned. In the second study Perlmutter compared individual and group recall of meaningful material. Groups recalled somewhat more correct information than individuals. However, the difference was not significant. On the other hand, it was found that individuals took significantly less time to recall the information than did the groups. In analyzing the content Perlmutter found very little correct information in the group's product which was not in the individual member's product, but he did find that some correct content which was in the individual product never found its way into the group product. While the group may be effective in elimination of certain incorrect information as was found by the earlier Perlmutter and Demontmollin study and also Shaw's, the group may also eliminate correct information.

In both of the above studies the critical control for pooling is also absent. There is no indication from the studies as to how well a group of subjects who worked separately and independently might perform. This is necessary to attempt to realize just what is happening in the interacting group.

Taylor and Faust (1952) compared individuals and groups of 2 and 4 using the old game of 20 questions as the task. Groups and individuals worked on 4 problems a day for 4 days. They found that groups of four were superior to individuals in terms of the number of questions asked, elapsed time to solution, and number of failures. They considered a notion mentioned earlier--the superior man idea--which says that a group's product is equal to the product of the best man in the group. While Watson's study did not support this notion, Taylor and Faust found that groups of 4 were about the same as the best individual of the day with respect to the number of questions to reach the solution. However, the best-man of the individuals was different each day. Taylor and Faust also neglected the crucial control for pooling, but the best man of the day may be a little similar to the best answer to each problem--pooling--as was mentioned earlier.

Faust (1959) in a later study used nominal groups against which he compared real groups. He gave individuals and groups spatial and verbal problems. From the individuals working independently on the problems he constructed nominal groups, i.e., he randomly placed four individual results together. If any of the 4 solved the problem within the given time limit, he counted it as a solution for the nominal group. The real groups were face to face groups working and discussing the problems. The spatial problems were "eureka" problems and his verbal problems were anagrams. He performed his experiments twice with the following results. In the first experiment the real groups were significantly superior to the nominal groups on the verbal problems but there were no significant differences on the spatial problems. In the repetition of the experiment he found no significant differences between the real and nominal groups. Faust recommends the use of his nominal groups as a control for the study of individual and group comparisons. Here is found the first recommendation for the use of a control for pooling although Faust does not label it as such. However, he does suggest that if a group has any facilitative effect it should produce something greater than that produced by nominal groups.

Marquart (1955) constructed groups working as individuals which were similar to Faust's nominal groups and compared their results with real group results on the Shaw problems. She found no significant differences between the two. By using the nominal groups--an experimental pooling--Shaw's results can be explained.

Taylor, Berry and Block (1958) studied the effectiveness of "brainstorming". Brainstorming involves a group consideration of a problem in an atmosphere which is free of criticism. Taylor et al. used groups which had frequently participated in small group discussions in the classroom instead of using accidentally formed groups which are typical of laboratory studies. To these groups they gave problems which had been previously tested and found the most suitable for brainstorming according to the criteria set by an acknowledged authority. Groups of 4 were given these problems and were asked to find solutions. They were also instructed in the brainstorming technique which involved an emphasis on quantity and lack of criticism. The investigators also gave the problems to individuals. The groups were found to be significantly superior to individuals. From the individual results nominal groups were constructed and a comparison of these nominal groups with real groups demonstrated that the nominal groups were superior in the total number of ideas produced, number of unique ideas produced and the quality of ideas. Taylor et al. concluded that brainstorming inhibits rather than facilitates problem solving of this nature.

In the studies which have used nominal groups with which to compare real groups the focus has been superiority and efficiency and not primarily upon how the group product is derived from individual products. Those studies generally point out that the group is the same or less efficient than the same number of individuals working separately, that is, the same or more can be obtained from persons working independently on a task than if the same persons were to work together on the task. While this is supported by the data and an interesting consideration, the primary problem of this paper is how the group product is obtained from individual products. The studies in which the nominal groups perform the same as real groups suggest that the group product derives from a pooling of individual products without group interaction having either facilitative or inhibiting effects.

Various conclusions have been drawn by reviewers from an examination of the research on group and individual problem solving. One aspect which is generally agreed on is that the task which is presented the subjects has an important effect on the results (Duncan, 1959; Johnson, 1955; Kelly and Thibaut, 1954; and Lorge et al., 1958). How the individual products become group products, the efficiency of group performance, and the influence of the group on the individual all depend upon the particular task used in the research.

Most of the research with group and individual performance has made use of laboratory groups. Lorge et al. point out the danger in generalizing from results using these "ad hoc" groups. Performance in a group which has been brought together for the goals of the experimenter may be expected to differ from performance in a group which has been formed for the accomplishment of mutual goals. While this danger clearly exists, Johnson (1955) points to the dependability of the studies done in the laboratory where the relevant variables can be controlled.

Another general conclusion reached by reviewers is that of superiority of average group performance over average individual performance. However, just what may be considered superiority must be examined closely. When such things as average time to solution of a complex problem or the average number of errors per trial are considered, the group is superior to the individual (Shaw, 1932; Gurnee, 1937b; Perlmutter and DeMontmollin, 1952). But when considering efficiency in man-hours expended, individuals are superior to groups (Husband, 1940; Taylór and Faust, 1952). Also, when real groups are compared to nominal groups there may be no difference, the nominal groups may be superior, or in one case the real group was superior (Watson, 1928; Taylor et al., 1958; Marquart, 1955; Faust, 1959).

As Kelly and Thibaut (1954) have pointed out, the older analyses of group problem solving processes have considered the group process analogous to the individual problem solving process (Dashiell, 1935; Bales, 1950), that is, the group proceeds through various phases such as, orientation, evaluation, etc. as does the individual solver. The emphasis on this analogy has led to the comparison between average group performance and average individual performance which has to some extent led further to a confusion as to what may be occurring in groups during either problem solving or learning. This, to some degree, may have influenced Barnlund (1959) in concluding that the superiority of groups in solving problems was due to "psychological factors inherent" in group discussion. Also, in group recall an analogy to individual performance may have resulted in Perlmutter and DeMontmollin concluding that the superiority of the group was due to facilitative effects of interaction in the group.

Furthermore, this analogy may have led Kelly and Thibaut (1954) to emphasize the uniqueness of group products, unique in the sense that they are different from individual products. It might be well to look into what they have to suggest about group products, even though their understanding of pooling is different from the author's. They take up the matter of how individual products become group products, but what they consider the pooling of individual accomplishments is what Timmons (1942) considered the averaging influence. Kelly and Thibaut (1954, p. 741) have this to suggest.

> The question . . . arises as to what accounts for the unique properties of group solutions as compared with pooled individual solutions. There are two logical possibilities, not mutually exclusive, either or both of which can account for this uniqueness. As a result of the group problem-solving situation and the interaction process involved, (1) *the individual solutions available for pooling or combination differ from the individual solutions derived under conditions of independent problem solving,* and/or (2) *the individual solutions are combined or assembled in a manner not reproducible by simple averaging, use of majority vote, or similar methods.*

(The emphasis is theirs.) They amplify what they mean under (1) above by suggesting that a change in individual solutions may come about by "modifications produced by direct social influence." By this is meant that a member in a group may, under social pressure from other members, accept a solution of another member or may attend to a specific line of thought suggested by another member. This may force the first member into considering a kind of solution that he would not have considered had he been working independently on the problem. Another factor which Kelly and Thibaut feel may cause the modification of individual solutions is the "social context" of individual problem solving. They feel that simply working in a group without any direct pressure being applied may raise the motivation level of the individual members so that they will perform differently than they would working alone. Another alternative which they do not emphasize is that social interaction might *lower* the motivational level. One member might become bored with other members socializing and simply not participate in the problem solving. In (2) above Kelly and

Thibaut suggest that the process of translating individual products to group products may not follow any simple procedure.

Allport (1924) has discussed some of the earlier research which studied the effect on individual performance of having others working on the same task. His classic concept of "social facilitation" arose from his own and others' experiments in which a number of subjects worked on a variety of tasks individually but in each others presence. These, however, are not interacting, co-operative groups of the kind we have been discussing. He found that for simpler tasks, e.g., multiplication, there was an increase in the output of an individual when working in the presence of others working on the same task. Allport also reports an experiment by Meumann (1914). Meumann read aloud a list of words to children and it was their task to recall as many as they could immediately after his reading. He found social facilitation with younger children, but none with older children. Recently, Allport (1962) has presented his current theoretical notions on these "co-acting" groups. He suggests that there is an increase in the motivation of individuals who are working in co-acting groups because they are concerned lest they "fall behind" or "fail to measure up to their own or others expectations." Whether or not this same facilitative effect occurs on individuals when they are part of an interacting co-operative group cannot be ascertained from the experiments with co-acting groups, but these experiments are of interest in understanding what effect just having other individuals around might have on a person. A direct determination of a facilitative effect in a co-operative rather than co-acting group is difficult because the group product is a result of the efforts of more than one individual and no measures of pure individual products are usually available. An exception to this is found in Zajonc's (1962b) study. He instructed individuals to work together for a group goal but also had measures of how each individual performed. The measure of group and individual performance was reaction time. Zajonc found that reaction time was facilitated in a co-operative group setting. However, it is difficult to know whether Zajonc actually had co-operative groups even though he instructed the group members to co-operate. This is because the group members received knowledge of the other members' performance, and the group setting may actually have been competitive rather than co-operative.

Kelly and Thibaut (1954) review Allport's earlier work and succeeding work of other researchers and have reached the following conclusions about working in the presence of others who are working on the same task or working with an audience:

> (a) Greater quantity of work where physical output is involved, suggesting increased motivation to perform the task.
> (b) Lesser quantity or quality of work where intellectual processes or concentration are involved, suggesting that social stimuli are able to compete successfully with the task stimuli.
> (c) Inhibitions of responses and qualitative changes in the work, which suggests that the person somehow "takes account" of the others as he goes about his work, e.g., he has fewer idiosyncratic thoughts, exercises moderation in judgment, and gives more popular or common associations.
> (d) Greater variations through time in his output, indicating the presence of periodic distractions and/or the effects of working under greater tension.
> (e) There is some evidence that these effects wear off as the person adapts to the social situation.

At another time Thibaut and Kelly (1959) take up facilitation and interference in the "dyadic" relationship. In discussing interference, which they emphasize, they suggest that one person's behavior may elicit behavior in another which is incompatible with the first person's thereby resulting in interference. Furthermore, both persons may interfere with each other. However, their conception of interference is somewhat vague. They suggest that interference may result in responses being performed less well, but also admit the possibility that "interference may not affect or may even improve the quality of performance." Here they are considering relatively simple and easy to perform tasks, and from their further elucidation they imply that the beneficial aspects of interference are caused by a member of a dyad changing sets which he would not do had he been performing independently. They also suggest that interaction may result in facilitation. This can occur when one member performs in such a way as to reward another person's behavior. This is reminiscent of Gurnee's (1939) social rewards.

From the preceding discussions it can be seen that certainty in the area of group products is not the case. Just how the group product is derived from the individual product is still a mystery. If it results from a pooling of individual accomplishments, first, the meaning of pooling is not agreed upon, and secondly, that which individuals produce in a group situation, the raw material for pooling, may be quite different than in an individual situation due to either facilitation or inhibition, or both. Another unanswered question is what is the cause of group superiority over individual performance when it exists? Is it simply pooling, is it pooling plus facilitation, or could it even be pooling plus inhibition? In the next chapter when the theoretical aspects are discussed more specifically it is hoped that the above questions will be somewhat clarified.

THEORY

Ekman (1955) and Lorge and Solomon (1955) have helped clarify the theoretical aspects of group products. Ekman differentiates four effects of co-operation in a group. The first effect which he mentions is "the individual effect" which says that an individual's behavior may be either facilitated or inhibited by the presence of others. This is similar to Kelly and Thibaut's notion which was mentioned earlier. The second effect, Ekman labelled as "the summation effect" and in this case the group product is merely the sum of the individual products. Watson's (1928) and Anderson's (1961) data were suggestive of this effect. The third effect is that of averaging which relates to the studies of Sherif (1935) and Timmons (1942). Ekman indicates that this can happen in the following situations: "When school marks, obtained from different teachers, are combined. Or when the decision of a committee is a compromise (not a majority decision) of the members' individual opinions--or when the same kind of decision is taken in the family." This is similar to what Kelly and Thibaut (1954) have said is pooling. The fourth effect is labelled "the probability effect" which says that in problem solving a group has a greater probability of producing a solution than does one individual simply because there are more persons working on the problem. This was illustrated previously in the studies of problem solving which used experimental controls for pooling in the form of nominal groups. Ekman's four effects summarize for the most part the theoretical questions raised by the studies of group products as they are derived from individual products.

The "summation effect" and the "probability effect" can both be considered examples of pooling, that is, the individual abilities may be pooled in a group to produce a longer list of words than produced by one individual working separately, or the abilities of the individual members of a group may be pooled so that there is a greater probability of a group solving the problem than an individual. The difference between the "summation effect" and the "probability effect" has to do with the difference in the tasks--the former relating to tasks involving parts which can be added, and the latter involving problem solving. Shortly, when it is shown how a pooling model can apply to recall and to problem solving the similarity between the summation and probability effects will be more clear.

In reaction to Shaw's experiment (1932, described earlier) Lorge and Solomon (1955) developed a mathematical model of group problem solving which was quite similar to Ekman's fourth effect, the probability effect. Like Ekman, their model states that the probability that at least one of the group members arriving at a solution is greater than the probability of success of one individual. The probability is greater because of the sheer number of individuals being lumped together and not because of any facilitative effects of the group situation. Let P_i be the probability that an individual will solve the problem within a given time. The $(1 - P_i)$ is the probability that an individual will not arrive at the solution. If k individuals are working on the problem together, but each has the probability, P_i, of solving and the probabilities are independent, then the probability that none of them solves is $(1 - P_i)^k$. Therefore, the probability of the group arriving at the solution, P_g, is equal to $1 - (1 - P_i)^k$. This model suggests no facilitative aspects of the groups, but merely represents the pooling of ability of the members. They applied their model to the Shaw data and, in most cases, were able to explain the group's superiority as due to the pooling of individual abilities. It appears as if Marquart and Faust in their construction of "nominal" groups did experimentally what Lorge and Solomon do mathematically.

From Lorge and Solomon's particular application one can suppose a general explanation for group superiority, that of a pooling of abilities. While it is not supposed that this rather simple notion can account for all group products, it seems a sensible place to start. Restle (1961) suggests the use of a pooling of abilities notion as an aid in identification of facilitative or interfering factors if they exist. These kinds of variables can be introduced "only as they are shown to push group attainment away from what would otherwise be predicted" on the basis of a pooling of abilities. Davis (1961) found that groups did not perform as well as was predicted by the Lorge and Solomon model. Davis used problems which involved stages to solution, and he found that what seemed to be occurring during group problem solving was that certain members would arrive at a solution which was incorrect but was satisfying enough for the member and the rest of the group so that they spent some time on this wrong answer. Since Davis' results were in terms of time to solution, this threw off the prediction based on a pooling model. However, he was able to take into account, mathematically, this interference and predicted the group results.

The first experiment of this thesis tests the Lorge and Solomon model in a group and individual recall setting similar to that of Perlmutter and DeMontmollin. Our reason for doing this replication is that Perlmutter and DeMontmollin did not present enough data to allow for a precise test of the pooling model, they did not test a sufficiently large number of groups, and from their description of their procedure, it was difficult to determine exactly what their method was. As was mentioned before, they gave a list of nonsense words to groups and individuals for five trials and the task was to recall as many words as possible at the end of each trial. The groups were superior throughout the five trials, and Perlmutter and DeMontmollin explained this superiority as being due to the facilitative effects of group discussion. Our first experiment differs from Perlmutter and DeMontmollin's in that (a) the number of words to be recalled has been reduced from 19 to 8; (b) both individual and groups have been given 9 instead of 5 trials; (c) more individuals and groups have been tested than did Perlmutter and DeMontmollin; (d) groups of 2 members and 3 members have been tested instead of just using 3 member groups.

The second part of the thesis involves an experiment studying group and individual solutions to the Denny Doodlebug problem (Rokeach, 1960). The use of a pooling of abilities notion may be sufficient to explain the results of group superiority in a simple task such as the learning of nonsense words, but it might be said that the task is such that rather complex cognitive functions are not involved and facilitation or interference due to interaction in a group did not get a chance to be demonstrated, whereas, in a more complex task they can be demonstrated. Using the Doodlebug problem allows for a test of the model in a more complex situation. A further advantage of the Doodlebug problem is that it allows for a separation of two phases of problem solving, analysis and synthesis (Rokeach, 1960). If there is evidence for either facilitation or interference in the problem solving as a whole, being able to isolate the two phases enables us to locate just where the facilitation or interference is occurring. We will be able to determine whether or not facilitation or interference is occurring in analysis or synthesis or both. This will provide us with more information to explain the results of the problem solving process as a whole.

Considering group recall (the first experiment) the notion of pooling of abilities might apply to Perlmutter and DeMontmollin's results in a way similar to Lorge and Solomon's application to Shaw's results. While the group remembered more nonsense words on each trial than did the individual, the group results may have been a pooling of words remembered by each of the members, and not attributable to any further facilitation from the group (a kind of summation effect).

To apply the Lorge and Solomon model to group recall, the following consideration must be made. Treat the probability of recalling a previously seen nonsense word within a given time just as Lorge and Solomon treat the probability of solving a problem within a given time. Let P_i = the probability that an individual will recall a given word. Then the probability of a group, P_g, of size k of recalling the word would equal $1 - (1 - P_i)^k$.

To predict Perlmutter and DeMontmollin's results directly from this model it would be necessary to know the individual and group results of recalling each of the words in the list, but these data were not available. However, a prediction of their results is possible if it is assumed that each of the words in the list has an equal probability of being remembered. Making this assumption, the proportion of the entire list recalled by the group, P_{tg}, can be estimated from the proportion of the entire list recalled by individuals, P_{ti}, by the following formula: $P_{tg} = 1 - (1 - P_{ti})^k$. It follows that the predicted group mean, $\bar{G}$, is found by multiplying P_{tg} by the number of words on the list.

Using the formula mentioned above, Perlmutter and DeMontmollin's group results were predicted. Only a part of their results was used since they had the following balanced design: two groups of subjects were tested both as groups and individuals. Group A was given 5 recall trials as individuals on a list of 19 nonsense words, paralogs, and then 5 recall trials as groups on a different, but comparable list. Group B was given a list to recall as a group followed by another to recall as individuals. The individual results of Group A were used to predict the group results of Group B. Table 2.1 describes the predicted and obtained means for each trial. It can be seen that the obtained and predicted results are quite close.

The Doodlebug problem involves an insect named Joe Doodlebug who operates under some rather strange conditions. Subjects given the

Table 2.1. Obtained and Predicted Means of Perlmutter and DeMontmollin

Trial	Individual means	Obtained group means	Predicted group means
1.	2.3	5.5	6.1
2.	5.4	10.2	12.0
3.	7.7	14.8	15.0
4.	9.7	17.3	16.8
5.	11.4	18.2	17.8

Doodlebug problem are required to describe how Joe can get to some food placed near him. There are a number of things which Joe can and cannot do, for example, he may jump in only four different directions, north, south, east and west; once he starts in any direction he must jump four times in that direction before he can switch to another direction; and various other conditions. Rokeach indicates that there are three beliefs which the subject must overcome before he can solve the problem:

> 1. The facing belief. In everyday life we have to face the food we are about to eat. But Joe does not have to face the food in order to eat it. He can land on top of it.
>
> 2. The direction belief. In everyday life we can change direction at will. But Joe is not able to do so because he is forever trapped facing north. Thus, the only way Joe can change direction is by jumping sideways and backwards.
>
> 3. The movement belief. When we wish to change direction in everyday life there is nothing to stop us from doing so immediately. But Joe's freedom of movement is restricted by the fact that once he moves in a particular direction--north, south, east or west--he has to continue four times in this direction before he can change it. Thus, when Joe stops. . . . he may have stopped in the middle of a sequence of jumps. . . . Many subjects have difficulty because they assume that Joe is at the end rather than possibly in the middle of a sequence.

Rokeach describes the phases in the solution to the problem. First, there is the analytic phase which involves changing each of the three beliefs mentioned above. A subject must discard his currently held beliefs and replace them with the new ones. This, however, is not enough to solve the problem for then the subject must integrate the three beliefs into a new system. This phase of the solution is considered the synthesizing phase. The subject is encouraged to think out loud while he is working on the problem and from his spontaneous comments and his questions to the experimenter it is possible to determine the time taken to overcome each of the three beliefs. These three times can be used as an indication of the subject's analyzing ability.

Rokeach (1960) mentions that it is difficult to determine exactly when the synthesizing phase of the solution begins. The subject must have *something* before he can start to synthesize it. It is likely that both the analysis and synthesis phase overlap one another, but since the total time to solution is known, it is possible to determine three synthesizing times by subtraction. The time to solve the problem after each of the three beliefs are overcome can be used as three measures of synthesis.

Using this problem in the study of group and individual problem solving allows for an isolation of the processes of analysis and synthesis by groups and individuals. The pooling of abilities model can, then, be applied to each of these two phases of problem solving. The Lorge and Solomon model when applied to Shaw's problems predicted the proportion of groups solving a problem within the time limit of the problem $P_g = 1 - (1 - P_i)^k$. Similarly, the proportion of groups solving within, say 5, 10, 15, 20, 25, 30, 35, and 40 minutes can be predicted. And, in a like fashion, the proportion of groups overcoming one belief within 5, 10, etc. minutes can be predicted. Each of the beliefs can be examined in this way, and this is a prediction of analysis by groups based on the analysis by individuals.

The prediction of synthesis by groups can be done in a similar fashion as the prediction of groups solving by the end of each of the times mentioned above. However, instead of predicting the proportion of groups who solved within 5 minutes from the start of the problem, the proportion of groups who solved within 5 minutes from the time they overcame a belief can be predicted. This prediction is based on the proportion of individuals who solved the problem in the first 5 minutes after overcoming a belief--using the same formula mentioned earlier. Predictions using each of the three measures of synthesis can be made.

It must be noted that the Doodlebug problem is a very difficult problem, and if subjects attempt to solve without being able to question an experimenter and without an experimenter repelling wrong solutions a very low percentage are successful. Marr (1961) reports that only .5 percent are able to solve under these conditions. In giving the Doodlebug problem to groups and individuals an experimenter must be present. This is a limiting factor in the study of group problem solving because the group is not allowed to operate in as free a situation as would be desirable. Many of the other studies did not have this limitation (e.g. Shaw, 1932; Marquart, 1955; Faust, 1959; Davis, 1961). However, the advantage of being able to study the processes of analysis and synthesis makes using the Doodlebug problem worthwhile, so long as we are aware of what the presence of an experimenter might produce. In Davis' study (1961) he found that the groups did more poorly than a pooling of abilities model predicted, and it appeared to be due to a group member misleading the group for sometime with a wrong answer. Restle and Davis (1961) have pointed out that if an experimenter is present to reject wrong answers, this kind of interference is unlikely to occur and the predictions based on a pooling of abilities model are likely to be more accurate. Because interference is kept at a minimum by this technique, facilitation, if present, would be more likely to evidence itself.

If facilitation does occur, using the Doodlebug problem enables a determination of where the facilitation occurred. If groups do better than predicted by the pooling of abilities model, the next question which can be examined is whether this superiority is in analysis, synthesis or both. Similarly, if the groups do worse than the pooling model predicts, an examination of the two phases of the problem solving process will locate where the group performs worse than expected by the pooling model.

To study the effect of group size, groups of 2 and 3 were tested. It would have been desirable also to have had groups of larger size. However, obtaining subjects for these groups presented a difficult problem. Sophomore college students have a tendency to forget some of their appointments and the chance of one subject in four or five not attending is fairly large so it was decided to use only groups of 2 and 3. Therefore, the conditions in both tasks were (a) 1 person present, (b) 2 persons present, and (c) 3 persons present which yields continuity of the increase in group size and allows for the study of the effect of group size. Taylor and Faust (1952) in their study, which was mentioned earlier, found groups of 4 significantly superior to groups of 2 and individuals; however, the groups of 2 were not significantly superior to individuals. Whether or not the lack of superiority of groups of 2 over individuals on Taylor and Faust's task was due to an insufficient sample, a sampling error, or was due to something particular to groups of 2 is unknown. In the two experiments of this thesis we will follow the adage that two heads are better than one and predict that as group size increases performance will be superior.

To briefly summarize the two experiments, experiment I involves giving groups of 2 and 3 and individuals a list of nonsense words to learn. The group learning will be predicted using the Lorge and Solomon pooling of abilities model. Experiment II involves giving a complex problem to groups of 2 and 3 and individuals. The problem used is one in which the analytic and synthesizing phases of problem solving can be measured. The group solutions, analysis and synthesis,will be predicted using the Lorge and Solomon pooling of abilities model. In both experiments the hypotheses are as follows. The larger the number of persons working on the tasks the better the performance will be. The other hypotheses are the null hypotheses, that is, the obtained group results are hypothesized as not differing significantly from group results which have been predicted on the basis of the Lorge and Solomon model. These hypotheses suggest that nothing either facilitating or interfering is occurring in the group process.

METHOD

EXPERIMENT I

Subjects

The Ss were 125 voluntary, introductory psychology and educational psychology students. Sixty-five of the Ss were tested in the fall of 1960 and the remaining 60 were tested in the summer of 1961. Fifty-one of the Ss were tested in 17 ad hoc groups of 3, 34 in 17 ad hoc groups of 2, and 40 individually. The Ss participated in the experiment by signing their name to a sign-up sheet which was either posted on a bulletin board near the introductory classes or passed around the classroom by the instructor. At each experimental time there were 3 spaces available for subjects to sign. If 3 subjects volunteered for and attended an experimental session, they were tested as a group of 3. If 2 Ss attended they were tested as a group of 2, and if 1 person attended he was tested individually. This type of assignment of Ss while not random did not contain any systematic bias and allowed for a complete utilization of the Ss who attended the experimental session. Forty-five males and 73 females served as Ss, but each group contained only members of the same sex.

Apparatus

Presentation board. A 20" x 4" x 2" board in which 8 slots were cut to hold the cards on which the nonsense words were written. A piece of 1/4" plywood was attached to the bottom of this board on which numerals were painted opposite the slots in the 20" x 4" x 2" board so that an easy selection of the cards was possible. At right angles to the 1/4" board was a 20" x 6" x 1" board on which there was a small ledge where the cards were placed by E during their presentation.

Recording device. A desk calender stand which contained a blank pad of paper and allowed for easy turning of the pages.

Procedure

The S or group of Ss sat at a table facing E and the presentation board. In front of S and in front of the groups on the table was a recording device. In 3 member groups the person who happened to sit in the middle was chosen by E as the recorder for the group, and in 2 member groups the person who sat to E's left was chosen as the recorder for the group.

The following instructions were given individuals by E: "The purpose of this experiment is to see if you can learn a list of 8 two-syllable nonsense words working alone. It is necessary that you make the maximum effort to remember the words as quickly as possible, the order with which you remember is not important. This is how we shall proceed. I will present a list a first time, each word written on a card at a regular rhythm. When all the words have been presented, I will say 'Go,' and you are to write on the paper before you the first word in the list that comes to your mind. It can occur that you remember a word but you are not certain of it, and think of another spelling. In this case, cross out the first word, and write below it the one you believe to be better until you reach a certain degree of agreement with yourself. Make certain that you flip the page before you write a new word. Only one word which is not crossed out per page is allowed. Do not turn the pages back once you have flipped them. You will be given 2 minutes to recall the words. At the end of this period I will say 'Okay,' and I will tell you to mark a large X on the next blank page and turn it over, and then we will proceed to the next trial and so on. Any questions?"

The following instructions were given the groups: "The purpose of this experiment is to find how you are going to learn a list of 8 two-syllable nonsense words, working as a group. Your individual results are not of interest, but the product of your collaboration is of primary importance. It is necessary to make the maximum effort to retain all the words as quickly as possible. The order in which you recall them is of no importance. This is how we shall proceed. I will present the list to you the first time, each word written on a card, presented separately and regularly. When all the words have been presented, I will say 'Go." If one of you remembers a word he tells the person who is sitting in the center (or on the right in the case of 2 member groups) and he writes the word on the paper in front of you, and the other person(s) as well as the one who gave the original should discuss this and other possible forms of the word until accord is reached. You should express your opinion as to the exact form of the word. Only one word may be adopted to represent the group. Cross out the words which you do not want. Unanimous agreement is not necessary. It is necessary that you all contribute; the order with which you intervene does not matter. When you have reached a decision as a group, flip the page on which that word is written before passing on to the next word. Once you have turned the page do not go back. You will be given two minutes in which to recall the words. At the end of that time I will say 'Okay,' and will tell you to mark a large X on the next blank page and turn it over, and then we will proceed to the next trial and so on. Any questions?"

Occasionally a group asked if they could divide the task with each member taking a certain proportion of the words. To this E replied, "No," and that this would give results of 3 individuals instead of group results, and the experiment was concerned with group results. E observed no apparent division of the task.

After the instructions were read, both individuals and groups were presented the stimuli in prearranged random orders. The stimuli, constructed by combining 3 letter nonsense syllables from Hilgard's (1951) list, were: BEMAW, HETIG, JALOB, KUQOM, LIPUF, MUWIY, SOHUJ, and ZADEC.

Groups and individuals were given 9 trials with each of the above words presented once each trial. The words were presented in a different random order each trial and, furthermore, the trial by trial order of the orders was randomized across individuals and groups. Each word was presented for 5 seconds followed by a 5 second interval in which no word was present. After the last word had been presented the Ss were given 2 minutes in which to recall the words. At the end of the ninth trial the Ss were told that the experimental session was concluded and the research was briefly explained.

EXPERIMENT II

Subjects

The Ss were 146 voluntary introductory psychology and educational psychology students, 80 of whom were tested in the winter of 1960-61 and the remaining 66 were tested in the summer of 1961. Forty-one subjects were tested as individuals, 48 in 24 groups of 2 and 57 in 19 groups of 3. The Ss volunteered for this experiment in the same way as they did in Experiment I, and, also, like Experiment I, they were tested as individuals, groups of 2 or 3 depending upon the number of subjects who attended the session. Sixty-six males and 80 females served as subjects, but each group contained only members of the same sex.

The Ss entered the experimental room and sat at a table facing E. The table was bare except for scratch paper.

The individuals were given the following instructions by E: "This is an experiment in problem solving. You'll be given one problem to solve. The problem is not a simple one but the solution can be reached by good logical analysis. Here is the problem. Read it over carefully."

At this point a copy of the Doodlebug problem (see the next page) was given the Ss.

When the individual indicated he had finished reading the problem, he was told by E: "I'd like to ask you to think out loud as you work the problem so I can let you know whether you are correct or not. You may ask me questions as you go along and you may refer to the problem at any time. You may use the scratch paper in any way you wish. Now, let's read the problem over together." E read aloud the problem while S followed on his copy of the problem. This was followed by E saying that they could begin.

The groups of 2 and 3 were given the following instructions: "This is an experiment in group problem solving. You'll be given one problem to solve. The problem is not a simple one but the solution can be reached by good logical analysis. Here is the problem. Read it over carefully."

At this point the groups were given the same problem as individuals and when they indicated they had finished reading the problem, E proceeded: "You are to work on the problem together and arrive at a group solution. You may talk as much as you want. In fact, it would be to your advantage to discuss the problem and your ideas on its solution. You may ask me questions as you go along and you may refer to the problem at any time. You may use the scratch paper in any way you wish. Now let's read the problem over together." E read aloud the problem as he did with individuals and then the group was told to begin.

Both groups and individuals were given 40 minutes to solve the problem and E recorded the time each belief was overcome and who in the groups overcame it. At the end of 40 minutes the individual or group was stopped if they had not solved and the research was explained and then the Ss were dismissed.

THE DOODLEBUG PROBLEM

THE CONDITIONS:

Joe Doodlebug is a strange sort of imaginary bug. He can and cannot do the following things:

1. He can jump in only four different directions; north, south, east and west. He cannot jump diagonally (e.g. southwest, northwest, etc.).
2. Once he starts in any direction, that is north, south, east or west, he must jump four times in that same direction before he can switch to another direction.
3. He can only jump, not crawl, fly, or walk.
4. He can jump very large distances or very small distances, but not less than one inch per jump.
5. Joe cannot turn around.

THE SITUATION:

Joe has been jumping all over the place getting some exercise when his master places a pile of food three feet directly west of him. Joe notices that the pile of food is a little larger than he. As soon as Joe sees all this food he stops dead in his tracks facing north. After all his exercise Joe is very hungry and wants to get to the food as quickly as he possibly can. Joe examines the situation and then says, "Darn it, I'll have to jump four times to get the food."

THE PROBLEM:

Joe Doodlebug was a smart bug and he was dead right in his conclusion. Why do you suppose Joe Doddlebug had to take four jumps no more and no less, to reach the food?

During the solution of the problem E was present and answered questions. E answered questions by referring to the appropriate conditions or statement in the problem when it was applicable which was the case for the majority of the questions. Wrong solutions to the problem were rejected also by referring to the appropriate part of the problem. At times S would ask a question which concerned one of the beliefs involved in the problem, such as, "Can Joe jump sideways?" In this case E replied in the affirmative, e.g., "Yes, Joe can jump sideways as well as backwards and forwards" and recorded the time when the question was asked as the time of overcoming the belief. Other questions were of a procedural sort such as, "Is Joe on a flat surface?" or "Is there a wall between Joe and the food?" which were answered appropriately by E.

The procedure followed in this experiment differed somewhat from the usual administration of the Doodlebug problem (see Rokeach, 1960) in which hints are given at variously spaced intervals. The hints are statements which allow the S to overcome each of the beliefs if he has not already done so, e.g. "Joe can jump sideways as well as backwards and forwards." These hints were not given in the present experiment because it would have involved an arbitrary cut off of the analysis process. Even though the analysis process could be studied up until the time the hints stop it, the analysis process involved in the solution of the problem would be arbitrarily stopped at the same time for groups and individuals, thereby confounding any facilitation or interference in relation to solution time. Another problem, however, is if the hints are not given, all the subjects will not overcome all the beliefs which limits the study of synthesis to only the Ss who overcame the beliefs.

RESULTS

Experiment I

Obtained Recall Results.

The first results to be examined are the obtained differences between groups and between groups and individuals. To be able to interpret the obtained mean differences properly the variances of the groups and individuals must first be considered. The variance for each condition for all 9 trials can be seen by referring to table 4.1.

Using the Hartley F-max test for homogeneity of variance (Walker and Lev, 1953) the variance of individuals was compared with that of groups of 3. The resulting F-ratio of 3.192, with $k = 3$ and n varying from 17 to 40, was significant, $p < .05$ for the smallest n. Therefore, the next step was to apply Bartlett's test for homogeneity of variance (Walker and Lev, 1953). This resulted in a $B = 9.052$, $df = 2$, $p < .02$. From these two tests the null hypothesis of homogeneity of variance can be rejected. A further comparison of the differences between each of the variances was made using the F-ratio test (Walker and Lev, 1953). These results are also in table 4.1.

The differences in the variance for all trials was either significant or approached significance at the .05 level and it was tempting to believe that the larger group size produced less variance, but a closer examination of the variances was made. This is presented in table 4.2. An inspection of this table indicates that the overall significant differences probably resulted from the differences in the later trials, and this is probably because the ceiling of 8 correct words per trial was approached soonest and more often by the groups of 3 and next by the groups of 2 resulting in the smallest variance in groups of 3 and the next smallest in groups of 2.

A F-max test (Walker and Lev, 1953) of the trial variances within each condition (individual, groups of 2 and groups of 3) was computed. These results demonstrate the existence of significant differences of variance within the group conditions (individual, F-max = 3.02, df = 39, k = 9, $p > .05$; groups of 2, F-max = 7.86, df = 16, k = 9, $p < .01$; groups of 3, F-max = 30.545, df = 16, k = 9, $p < .01$). In the group conditions the difference in variance again is likely due to the ceiling of 8 correct being reached by groups and not individuals.

The previous results indicated that an analysis of variance would be inappropriate due to the heterogeneity of variance. However, it was felt that an examination of the differences between the individual and groups by use of the *t*-test would also be inappropriate without a prior analysis of the data by an analysis of variance. Therefore, a suitable analysis of variance technique was used in order to determine if there were overall differences between the groups and individuals and whether or not there was a change over trials, keeping in mind that the data contained differences in variance.*

The kind of analysis of variance which was selected as appropriate for repeated measures on separate groups was Lindquist's (1953) type I design. The purpose of this design was to test the significance of the overall differences between individuals, groups of 2, and groups of 3; and also the overall difference in the recall trials, and finally to test the interaction between the trials and conditions (individuals, groups of 2 and groups of 3). The results of the analysis of variance are presented in table 4.3.

---

*The Norton Study (as reported by Lindquist, 1953) suggests that heterogeneity of variance is not of serious consequence so long as the level of significance is made more rigorous.

Table 4.1. Obtained recall results: A comparison of the overall variances of individuals, groups of 2 and groups of 3.

Condition	Variances
Individual	109.938
Groups of 2	41.941
Groups of 3	34.441

Comparison of the conditions

Comparison	F	df	P
Inviduals vs. groups of 2	2.621	39 and 16	< .01
Individuals vs. groups of 3	3.192	39 and 16	< .01
Groups of 2 vs. groups of 3	1.218	16 and 16	> .05

Table 4.2. Obtained recall results: Comparisons of obtained trial variances.

	Variances			F-ratios		
Trial	Ind.	Grp 2	Grp 3	Ind vs Grp 2	Ind vs Grp 3	Grp 2 vs Grp 3
1.	1.051	1.691	1.360	-1.609	-1.294	-1.243
2.	2.708	2.066	3.360	1.311	-1.241	-1.626
3.	3.179	2.721	2.279	1.168	1.395	1.194
4.	2.910	.934	.632	3.116**	4.604**	1.478
5.	2.767	.515	.346	5.373**	7.997**	1.488
6.	2.459	1.279	.154	1.922*	15.968**	8.305**
7.	1.833	.346	.110	5.298**	16.664**	3.145*
8.	2.256	.757	.279	2.980**	8.086**	2.713*
9.	1.269	1.140	.191	1.113	6.644**	5.968**

Note: The ratios which are preceded by a minus (-) sign are those which have as the denomenator the variance of individuals in the cases of comparisons with individual variances, and the denomenator is the variance of groups of 2 in the comparisons of groups of 2 with groups of 3.

$^{*}p < .05$.

$^{**}p < .01$.

Table 4.3. Analysis of variance of obtained recall data

Source	df	sum of squares	mean square	F-ratio
Between Ss	73	891.045		
Condition (Ind, grps)	2	278.853	139.426	16.171[*]
Error A	71	612.192	8.622	
Within Ss	592	2211.778		
Trials	8	1648.526	206.066	230.757[**]
Condition X Trials	16	56.006	3.500	3.919[**]
Error B	568	507.246	.893[***]	
Total	665			

[*] $p < .001$, error term A used for F-ratio.

[**] $p < .001$, error term B used for F-ratio.

[***] It is likely that the extreme smallness of this error term probably to a large measure, reflects the fact that several of the trials were performed when most groups and many individuals were performing perfectly, and thereby, reducing the variance. If this were not the case, the confusing result of the significant Condition X trials interaction might not exist.

The results of the analysis of variance indicated a difference in trials which supports the notion that there was learning during the task. Also, the results suggested a difference between the conditions. Therefore, individual t-tests were performed between individuals and groups of 2, individuals and groups of 3, and groups of 2 and groups of 3. The results of these tests are presented in table 4.4. The difference between individuals and both groups of 3 and 2 was significant. The difference between groups of 3 and 2 was not quite significant, however, this difference approached significance at the .05 level.

Obtained vs. predicted recall results.

The mean words correct for all trials for groups of 2 and groups of 3 was compared with the predicted mean words correct. The predicted mean total for groups of 2 and groups of 3 was computed by applying the Lorge and Solomon model to each trial. Using the proportion of words recalled correctly on each trial by individuals as the estimate of P_{ti}, the formula, $P_{tg} = 1 - (1 - P_{ti})^k$ was used to predict the proportion of words recalled correctly by groups on each trial. By multiplying P_{tg} by the number of words on the list (8) a predicted mean, $\overline{G}$, for each trial was calculated, and by summing the predicted means for the nine trials, a predicted mean for all trials was calculated. These predicted means of the total for groups of 2 and groups of 3 are compared with the obtained means of the total in table 4.5.

From the table it may be seen that for groups of 2 the prediction did not differ significantly from the obtained. But, the prediction for groups of 3 did differ significantly from the obtained. However, it must be pointed out that there is no adequate statistic available to test the closeness of the predictions. The t-test was used but in so doing the variance of the predictions was assumed to be zero and the predicted mean was treated as a parameter. The predictions, since they were

Table 4.4. T-test comparison of obtained recall results.

Comparison	Mean diff.	Std. error mean	t	p	df
Inds. vs. grps 2	9.365	2.284*	4.100	< .01	55
Inds. vs. grps 3	13.365	2.185*	6.117	< .01	55
Grps 2 vs. grps 3	4.000	2.120	1.887	> .05**	32

*The standard error of the mean was obtained by using a pooled method which takes into account the heterogeneity of variance (Edwards, 1956).

**A t of 2.042 was needed for p = .05.

Table 4.5. Recall, total means--predicted and obtained.

	Means			Std. error			
	Obtain.	Predict.	Diff.	of mean	df	t	p
Groups of 2	58.765	61.720	2.955	1.521	16	1.881	> .05*
Groups of 3	62.765	66.168	3.403	1.423	16	2.391	< .05*

* t = 2.12 for p = .05

based on the individual results, are variable; but there was no adequate estimate of this variance. The assumption of zero variance of the predictions was overly stringent for the purposes of this study. While it is possible that the difference between the obtained and predicted for groups of 3 would not be significant if the variance of the predictions were known, the difference is still large enough to suggest a flaw in the Lorge and Solomon model.

The obtained results for each trial was compared to the predicted results for each trial and this can be seen by referring to figures 4.2 and 4.3. From inspection it can be seen that, generally, the predicted and obtained are quite close, both for groups of 2 and groups of 3 except for the first trials. Tables 4.6 and 4.7 present the results of t-test comparisons of obtained and predicted results of groups of 2 and 3 for each trial. The trial by trial t-tests indicate that the difference between the obtained and predicted on the first trial for both groups of 2 and 3 is significant with the stipulation that the t-test is overly stringent as mentioned above.

The significant difference on the first trial may indicate a weakness in the assumption that all words are equally likely to be remembered. While the positions of the words were randomized on each trial, a serial position effect could still occur on the first trial. If so, the assumption that each word has an equal chance of being recalled is not tenable. Table 4.8 shows the proportion of correct recalls by individuals and groups at each of the eight serial positions, on trial 1. It can be seen that the first word presented was recalled by 60 per cent of the individuals while the words in the middle of the first presentation were recalled by less than 10 per cent of the individuals, indicating a strong serial position effect.

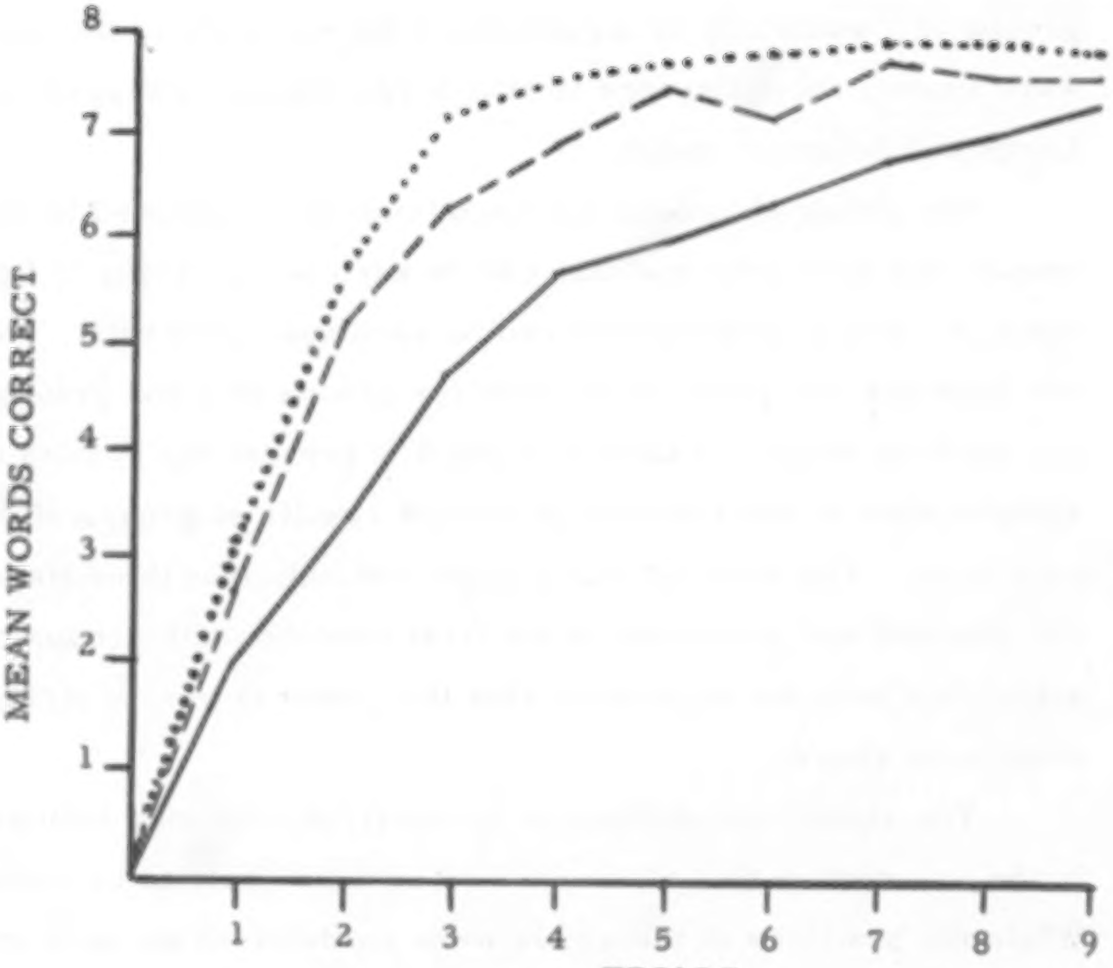

Figure 4.1. Obtained recall results. Groups of 2 (broken line), groups of 3 (dotted line) and individuals (solid line).

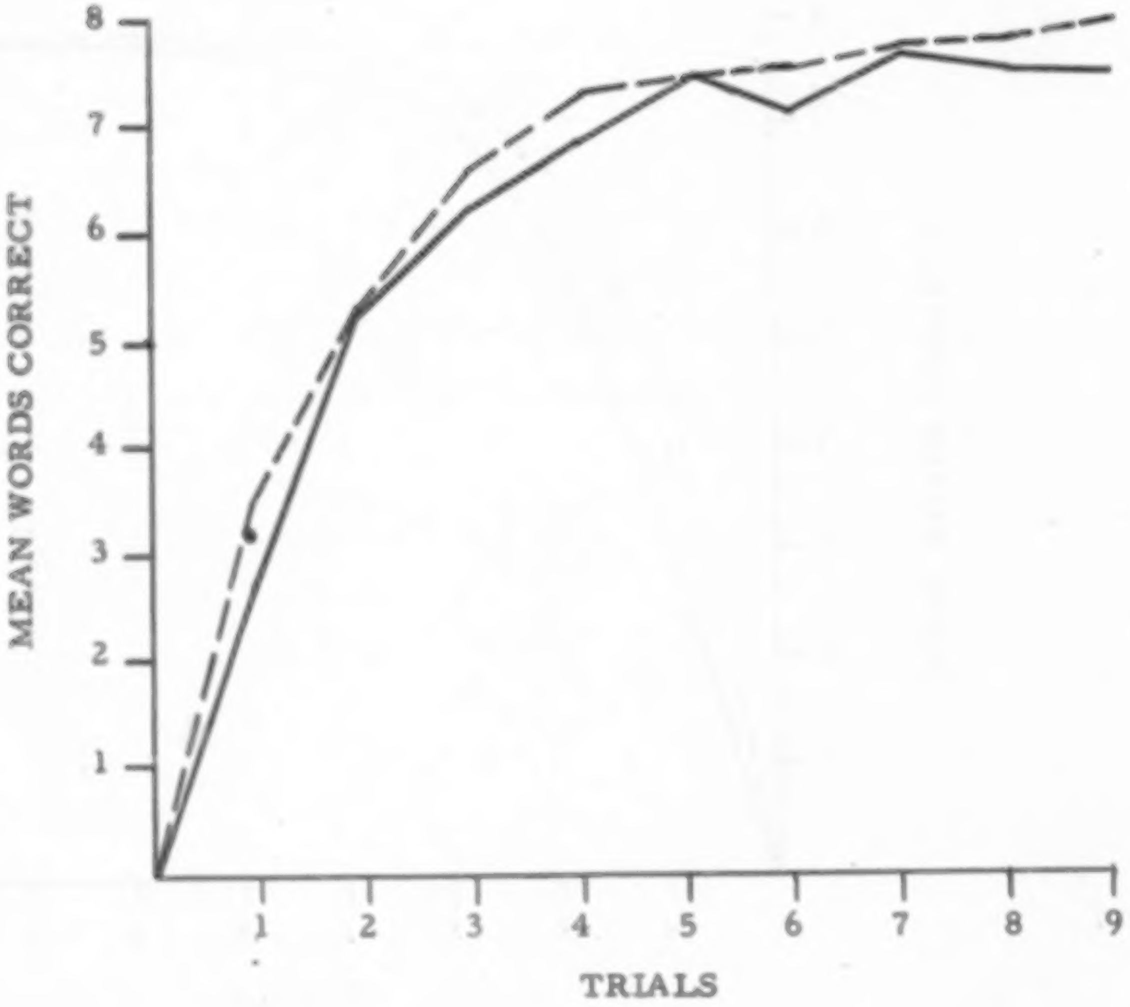

Figure 4.2. Obtained (solid line) and predicted (broken line) recall results of groups of 2. The dot is the corrected first trial prediction.

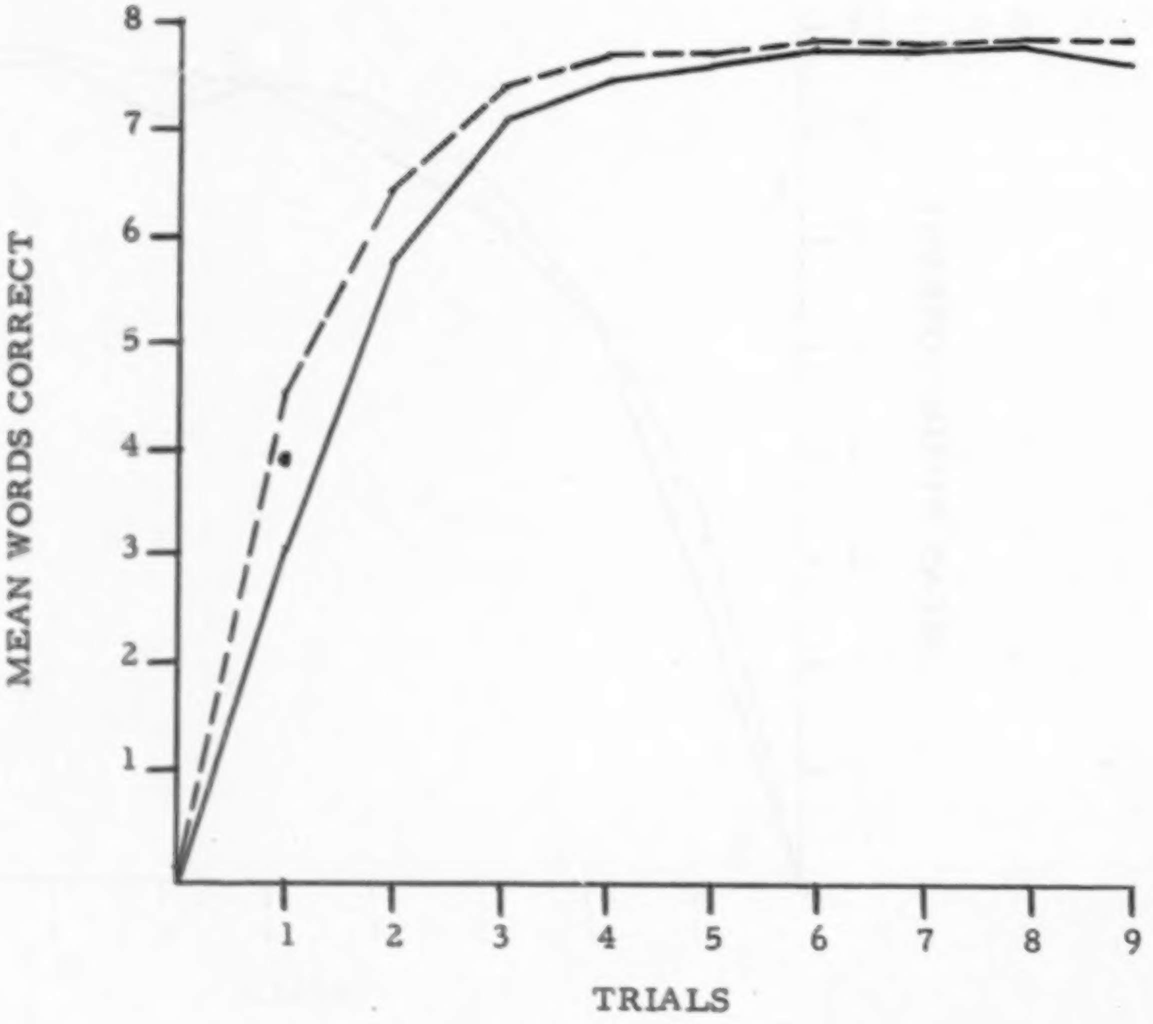

Figure 4.3. Obtained (solid line) and predicted (broken line) recall results of groups of 3. The dot is the corrected first trial prediction.

Table 4.6. Obtained vs. predicted recall results trial by trial for groups of 2.

Trial	Mean difference	Std. error of mean	t
1.	.760	.3153	2.410*
2.	.112	.3485	.321
3.	.368	.3990	.920
4.	.424	.2342	1.810
5.	.008	.1739	.046
6.	.512	.2743	1.866
7.	.096	.1426	.067
8.	.280	.2110	1.327
9.	.400	.2590	1.544

*$p < .05$

Table 4.7. Obtained vs. predicted recall results trial by trial for groups of 3.

Trial	Mean difference	Std. error of mean	t
1.	1.528	.2828	5.403*
2.	.600	.4455	1.350
3.	.264	.3662	.721
4.	.232	.1929	1.203
5.	.168	.1426	1.178
6.	.120	.0953	1.259
7.	.088	.0805	1.093
8.	.160	.1283	1.247
9.	.232	.1059	2.190*

*$p < .01$

Table 4.8. First trial proportions correct for individuals, groups of 2, and groups of 3 and predicted proportions correct for groups of 2 and groups of 3.

Position	Proportion of individuals	$1 - (1-Pi)^2$	Proportion of groups of 2	$1 - (1-Pi)^3$	Proportion of groups of 3
1	.600	.824	.840	.936	.647
2	.425	.669	.529	.810	.529
3	.175	.319	.353	.438	.353
4	.050	.098	.118	.143	.235
5	.075	.144	.235	.208	.235
6	.075	.144	.059	.208	.294
7	.225	.399	.235	.534	.118
8	.400	.640	.412	.784	.706
Sum ($\overline{G}$)		3.253		4.061	

What seemed to have happened on the first trial was that each of the group members tended to learn the same words. The pooling of abilities model with the assumption of each word having an equal chance of being recalled does not take this into account. The model pools the ability of all members with a disregard to which words were learned. An extreme example of this would be if all members of the groups and all individuals learned only the first and last words on the list. In this case the group would perform no better than the individual, but the prediction of the group's performance would be higher than the actual performance.

Because of the serial position effect, new predicted means, $\bar{G}$'s, for the first trial were made, based on the proportion of individuals recalling the words in each of the 8 positions. The formula, $P_g = 1 - (1 - P_i)^k$, was applied separately for the words of each position. For groups of 2 and groups of 3 the new prediction of the mean number of words recalled was made by summing the predicted proportions of the words recalled by the group in each position, $\bar{G} = \Sigma P_g$ (see table 4.8).

The new prediction of 3.25 correct responses on trial 1 for groups of 2 was found not to differ significantly from the obtained value of 2.76 ($\underline{t} = .847$, df = 16, $p > .05$). However, the new prediction of 4.06 for groups of 3 was found to differ significantly from the obtained value of 3.12 ($\underline{t} = 3.325$, df = 16, $p < .05$).

A new predicted mean words correct for all trials for groups of 3 was computed using the new first trial prediction, taking into account the serial position effect. The new predicted mean total of 65.581 was compared to the obtained of 62.765 and the difference of 2.816 was found not to be significant ($\underline{t} = 1.978$, df = 16, $p > .05$). Previously, the difference between the total predicted and obtained was significant (see above) when the serial position effect was not taken into account.

Generally, the predicted and obtained results are quite close. However, the still significant difference between the predicted and obtained performance of groups of 3 on the first trial after the serial position effect is taken into account is unexplained. Also unexplained is the difference between the predicted and obtained performance of groups of 3 on the last trial.

The obtained results indicated the expected group superiority, groups of 2 and 3 being superior to individuals, however, the difference between groups of 2 and 3 was not quite significant.

Experiment II

Obtained Doodlebug results.

The time to solution of the Doodlebug problem for groups of 2, groups of 3 and individuals was recorded. Not all individuals or groups of 2 solved the problem so it was impossible to compute mean solution times and variances of solution times. The measures used for comparisons of the groups and individuals were proportions.

The following measures were used to compare the obtained results of groups and individuals on the Doodlebug problem:

1. Solution. The proportion of groups of 2, groups of 3 and individuals who solved the problem in the 40 minute time limit were compared.

2. Analysis. The proportions of groups of 2, groups of 3 and individuals who overcame 1 belief and who overcame 2 beliefs were compared.

3. Synthesis. The proportions of groups of 2, groups of 3 and individuals who solved after overcoming 1 belief and after overcoming 2 beliefs were compared.

The obtained proportions of solutions for each condition are presented in table 4.9 along with the obtained proportions relevant to the analysis and synthesis measures.

Table 4.9. **Obtained Doodlebug results: Proportions of individuals, groups of 2, and groups of 3 passing solution, analysis, and synthesis measures.**

		Individuals	Groups of: 2	3
SOLUTION	**Proportion solving**	**.658**	**.667**	**1.000**
ANALYSIS	**Proportion overcoming 1 belief**	**1.000**	**1.000**	**1.000**
	Proportion overcoming 2 beliefs	**.854**	**.792**	**1.000**
SYNTHESIS	**Proportion solving after overcoming 1 belief**	**.658**	**.667**	**1.000**
	Proportion solving after overcoming 2 beliefs	**.771***[*]	**.842***[*]	**1.000***[*]

---

[*] **To compute these proportions the number of groups or individuals who overcame 2 beliefs was used as the denomenator. In the case of individuals this was 35 of the total of 41. For groups of 2, it was 19 of 24, and for groups of 3 it was 19 of 19.**

Before presenting further results, the disappearance of one of the beliefs must be noted. Looking at the tables 4.9, 4.10 etc. it can be seen that the measures of analysis and synthesis use only two beliefs, whereas, in the introduction it was mentioned that there were three beliefs involved in the solution of the Doodlebug problem. In the course of the administration of the Doodlebug problem in a setting without any hints it was found by E that it was difficult to determine when, if at all, the subjects overcame a particular one of the beliefs. The particular belief was the facing belief, i.e., "Joe does not have to face the food in order to eat it." Many subjects never asked a question about Joe's facing or not facing the food, but they still solved the problem, and it was evident in their solution of the problem that they had assumed that Joe did not have to face the food in order to eat it. This meant that either during their working on the problem they overcame this belief and when this occurred it was not obvious to E or that they started the problem assuming that the bug did not have to face the food. Because of the difficulty in determining when this particular belief was overcome, it was decided not to include this belief in the calculation of the analysis and synthesis measures.

Looking at table 4.10 comparing individuals and groups of 2 it can be seen that in no comparisons are groups of 2 significantly different from individuals. This can be further amplified by examining figures 4.4 through 4.8 which show the close similarity between individuals and groups of 2 on the Doodlebug problem.

Table 4.11 compares individuals and groups of 3. An examination of this table reveals that individuals and groups of 3 differ significantly on 3 of the 4 measures of analysis and synthesis and also the total proportions solving the problem. This is also indicated by figures 4.4 through 4.8.

Table 4.10. Obtained Doodlebug results: Individuals vs. groups of 2, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis.*

		Difference of proportions	z	p
SOLUTION	solving	.009	.086	.93
ANALYSIS	overcoming 1 belief	(no difference, both proportions = 1.00)		
	overcoming 2 beliefs	.062	.607	.54
SYNTHESIS	solving after overcoming 1 belief	.009	.086	.93
	solving after overcoming 2 beliefs	.071	.624	.54

---

*The statistic used to test the difference between the proportions is given in Lorge and Solomon (1955) and uses the following formula:

$$z = \frac{\theta_1 - \theta_2}{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

Where $\theta = 2 \arcsin \sqrt{\text{proportion solving}}$, N = sample size.

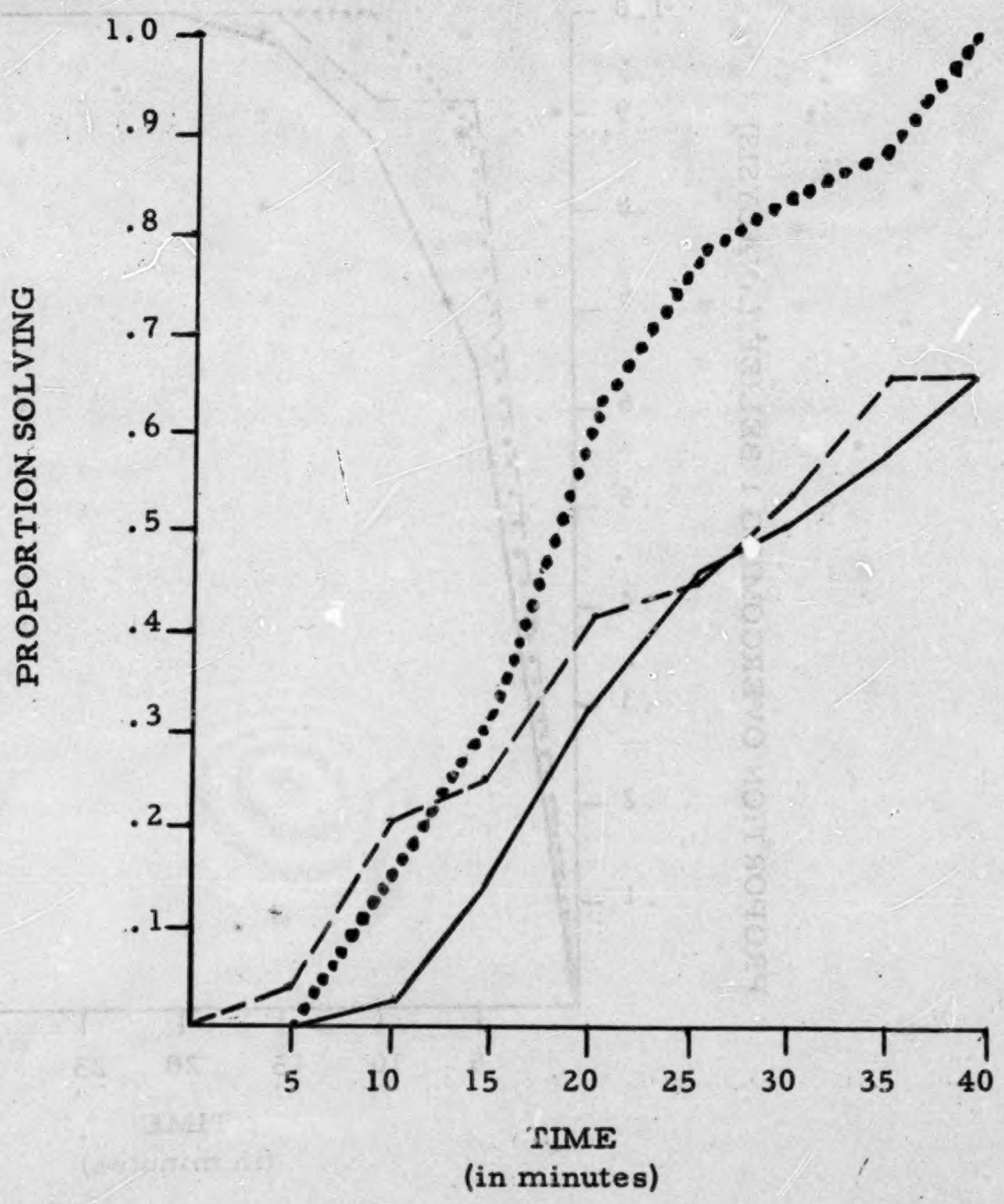

Figure 4.4 Obtained Doodlebug results. Cumulative distributions of proportions of solutions in groups of 2 (broken line), groups of 3 (dotted line) and individual (solid line) conditions.

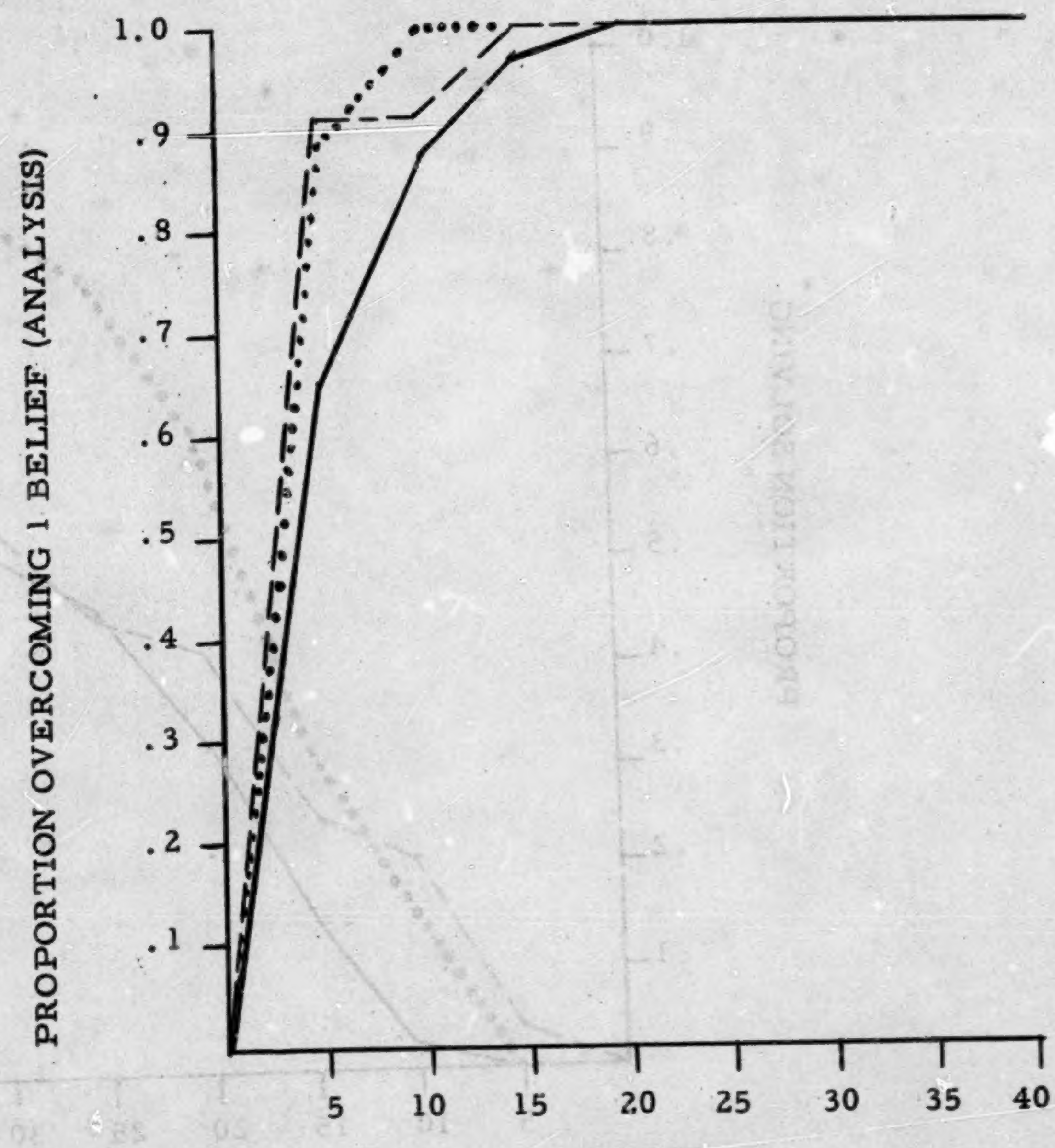

Figure 4.5. Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2 (broken line), groups of 3 (dotted line) and individuals (solid line) overcoming 1 belief.

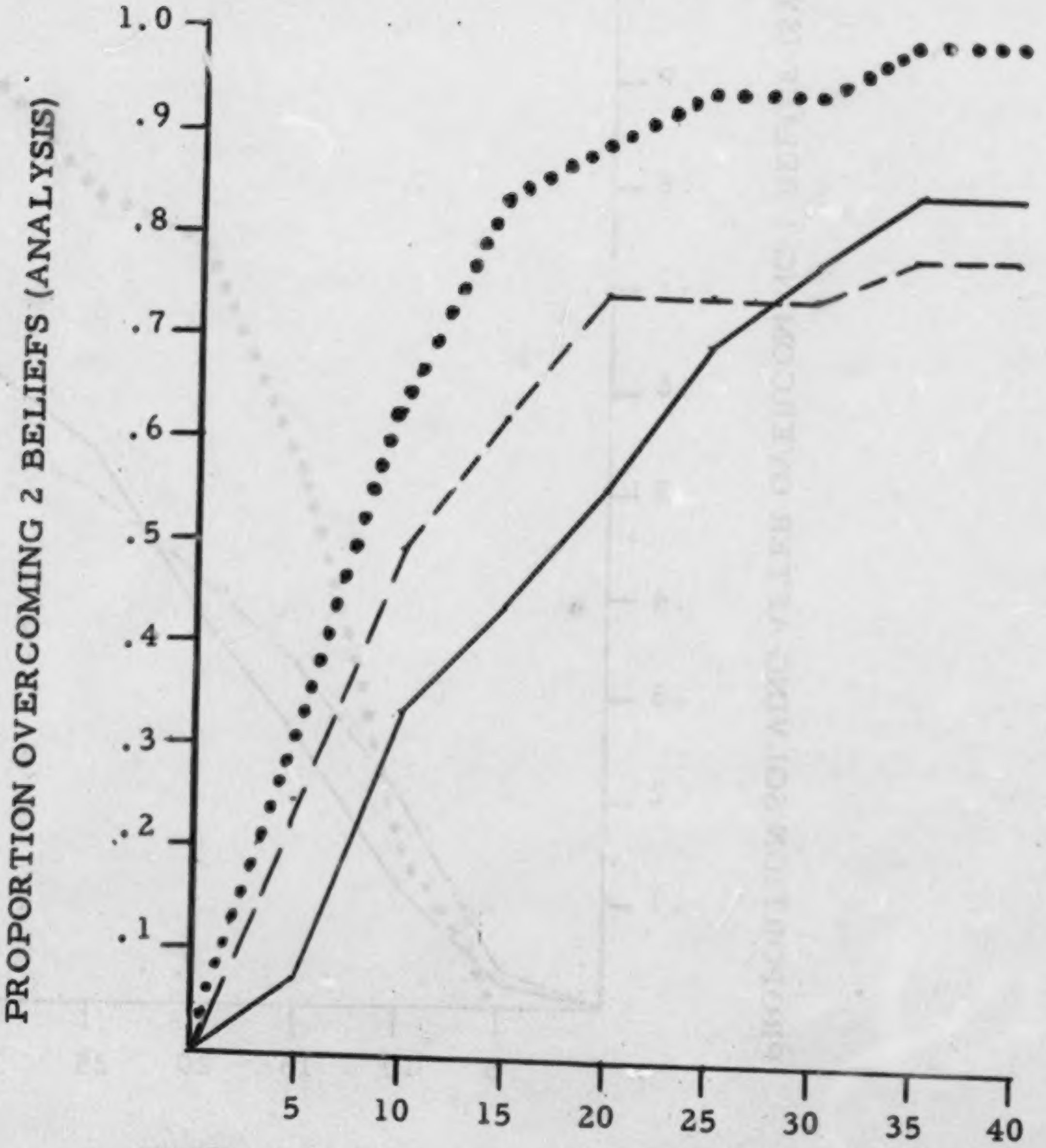

Figure 4.6. Obtained Doodlebug Results. Cumulative distributions of proportions of groups of 2 (broken line), groups of 3 (dotted line) and individuals (solid line) overcoming 2 beliefs.

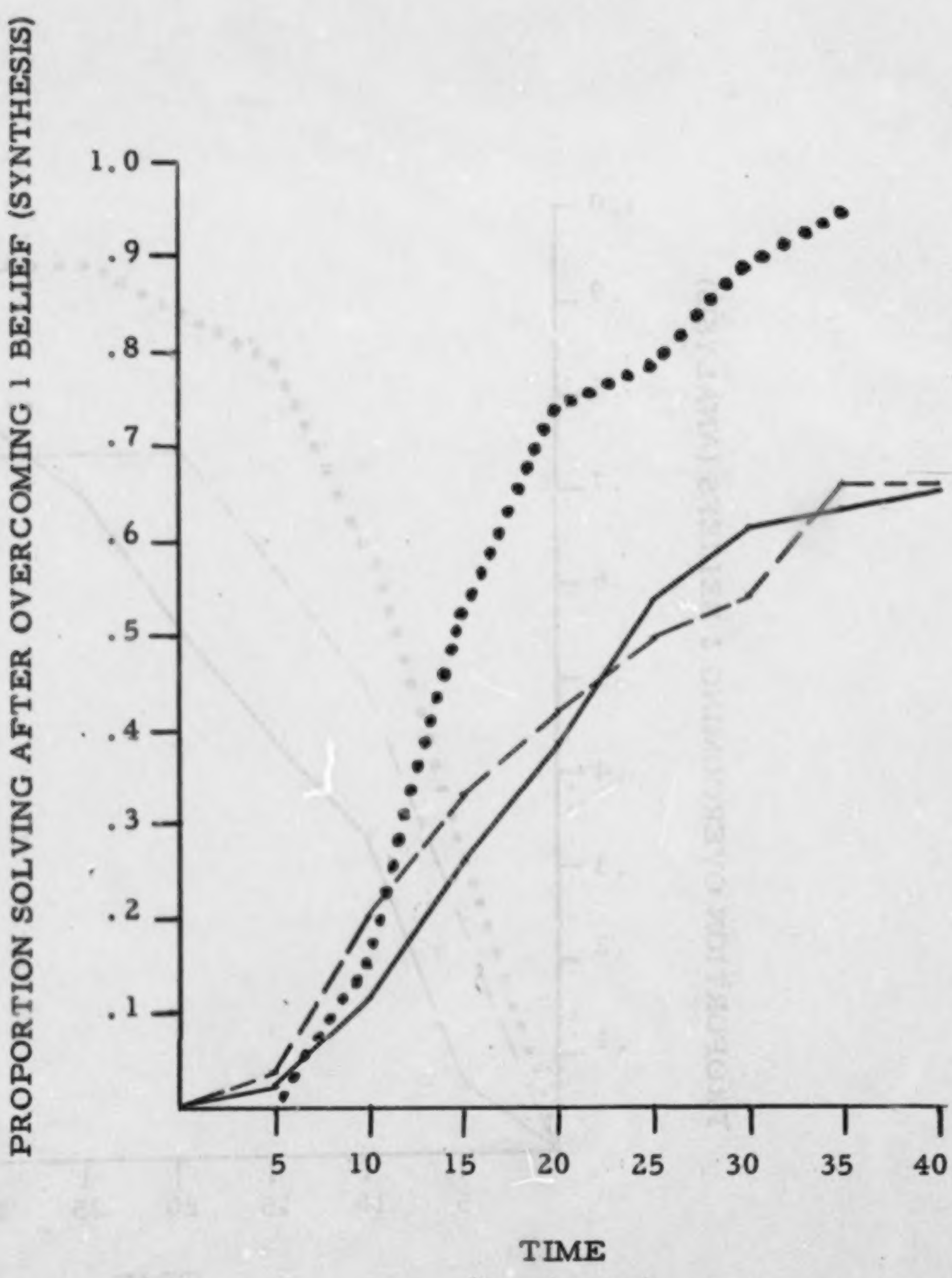

Figure 4.7. Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2 (broken line), groups of 3 (dotted line), and individuals (solid line), solving after overcoming 1 belief.

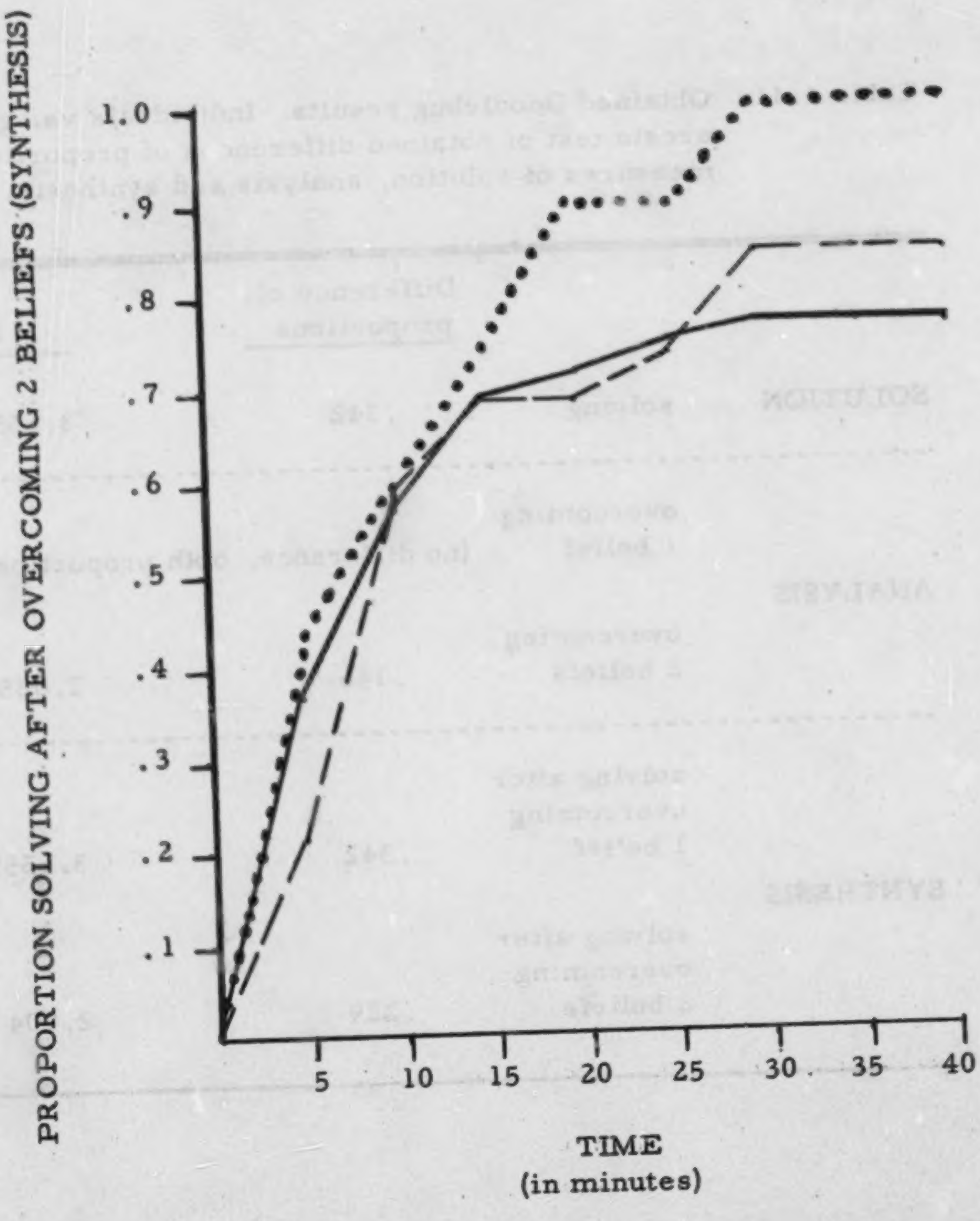

Figure 4.8. Obtained Doodlebug results. Cumulative distributions of proportions of groups of 2 (broken line), groups of 3 (dotted line) and individuals (solid line) solving after overcoming 2 beliefs.

Table 4.11. Obtained Doodlebug results: Individuals vs. groups of 3, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis.

		Difference of proportions	z	p
SOLUTION	solving	.342	3.655	< .001
ANALYSIS	overcoming 1 belief	(no difference, both proportions = 1.00)		
	overcoming 2 beliefs	.146	2.035	< .05
SYNTHESIS	solving after overcoming 1 belief	.342	3.655	< .001
	solving after overcoming 2 beliefs	.229	2.704	< .01

A comparison of groups of 2 and groups of 3 as given by table 4.12 yields differences. As the table indicates groups of 2 differ from groups of 3 on 3 of the 5 measures. The proportion of groups of 2 solving after overcoming 2 beliefs does not differ significantly from the proportion of groups of 3, but the level of significance of .07 is quite close to the usual cut-off level of .05. Since the majority of the comparisons yields significant differences, a conclusion that groups of 2 differ from groups of 3 in solving the Doodlebug problem can be made. This conclusion can be further supported by examining figures 4.4 through 4.8.

The group superiority which was found in the recall experiment is not the same in the Doodlebug experiment. Whereas groups of 3 are significantly superior to individuals, groups of 2 do not come close to being significantly different from individuals, and, as this suggests, groups of 3 are significantly superior to groups of 2.

Obtained vs. predicted Doodlebug results.

The proportions of groups of 2 and groups of 3 solving the Doodlebug problem were predicted by substituting the proportion of individuals solving the problem for P_i in the following formula, $P_g = 1 - (1 - P_i)^k$, where, as was previously mentioned, P_g = the predicted proportion of groups solving the problem and k = the number of persons in the group. The proportions passing the measures of analysis and synthesis (see page 51) were similarly predicted. The results of these predictions for groups of 2 are presented in table 4.13 along with the results of an arcsin test of the difference between the obtained and predicted proportions. These results for groups of 3 are presented in table 4.14. It must be noted that the arcsin test of the difference between predicted and obtained proportions is overly stringent for our purposes for the same reason the t-test was overly stringent for a test between obtained and predicted means in the recall experiment.

Table 4.12. Obtained Doodlebug results: Groups of 2 vs. groups of 3, arcsin test of obtained differences of proportions passing measures of solution, analysis and synthesis.

		Difference of proportions	z	p
SOLUTION	solving	.333	3.238	< .002
ANALYSIS	overcoming 1 belief	(no difference, both proportions = 1.00)		
	overcoming 2 beliefs	.208	2.351	< .02
SYNTHESIS	solving after overcoming 1 belief	.333	3.238	< .002
	solving after overcoming 2 beliefs	.158	1.829	.07

It can be seen from table 4.13 that the predictions of the performance of groups of 2 was not very accurate. Three of the 5 predictions differed significantly from the obtained results, and a fourth prediction differed at the .10 level of significance. In all of these predictions the group performed more poorly than the pooling model predicted.

In contrast to the predicted performance of groups of 2, it can be seen from table 4.14 that the predicted performance of groups of 3 is quite accurate. In no instance did the absolute difference between the predicted and obtained proportions exceed .05, and none of the differences approached the .05 level of significance. The obtained and predicted values are near or at the upper limit of 1.00. However, the values obtained by the arcsin transformation as presented by Walker and Lev (1953) are still reasonably sensitive at these extremes.

The above predicted and obtained proportions for groups of 3 are quite close, but the comparison concerns the proportions solving at the end of the 40 minute time limit and does not indicate whether or not the Lorge and Solomon model would be an adequate predictor throughout the 40 minute time interval. This is particularly relevant because the obtained and predicted proportions are at or near the limit of 1.00. It is reasonable to suppose that the obtained results could yield many groups of 3 solving in the last 5 or 10 minutes thus indicating a discrepancy between the obtained and predicted. To remedy this, cumulative distributions of proportions solving and passing the measures of synthesis and analysis at 5 minute intervals were constructed. Predicted proportions for each of the intervals were also constructed. Figures 4.9 through 4.18 present the comparison between the predicted and obtained distributions.

Observing figures 4.9 through 4.13 for groups of 2 it can be seen that the predicted and obtained generally differ. The results of the Kolmogrov-Smirnov test for groups of 2 is presented in table 4.15.

Table 4.13. Obtained vs. predicted Doodlebug results: Arcsin test of the difference between obtained and predicted proportions of groups of 2 solving the Doodlebug problem and passing measures of analysis and synthesis.

		Obtained proportion	Predicted proportion	Difference	z	p
SOLUTION	solving	.667	.883	.216	2.530	.011
ANALYSIS	overcoming 1 belief	1.000	1.000	.000	.000	-
	overcoming 2 beliefs	.792	.979	.187	3.239	.01
SYNTHESIS	solving after overcoming 1 belief	.667	.883	.216	2.530	.011
	solving after overcoming 2 beliefs	.842	.948	.106	1.622	.10

Table 4.14. Obtained vs. predicted Doodlebug results: Arcsin test of the difference between obtained and predicted proportions of groups of 3 solving the Doodlebug problem and passing measures of analysis and synthesis.

		Obtained proportion	Predicted proportion	Difference	z	p
SOLUTION	solving	1.000	.960	.040	.753	.45
ANALYSIS	overcoming 1 belief	1.000	1.000	.000	.000	-
	overcoming 2 beliefs	1.000	.997	.003	.524	.60
SYNTHESIS	solving after overcoming 1 belief	1.000	.960	.040	.753	.45
	solving after overcoming 2 beliefs	1.000	.988	.012	.045	.96

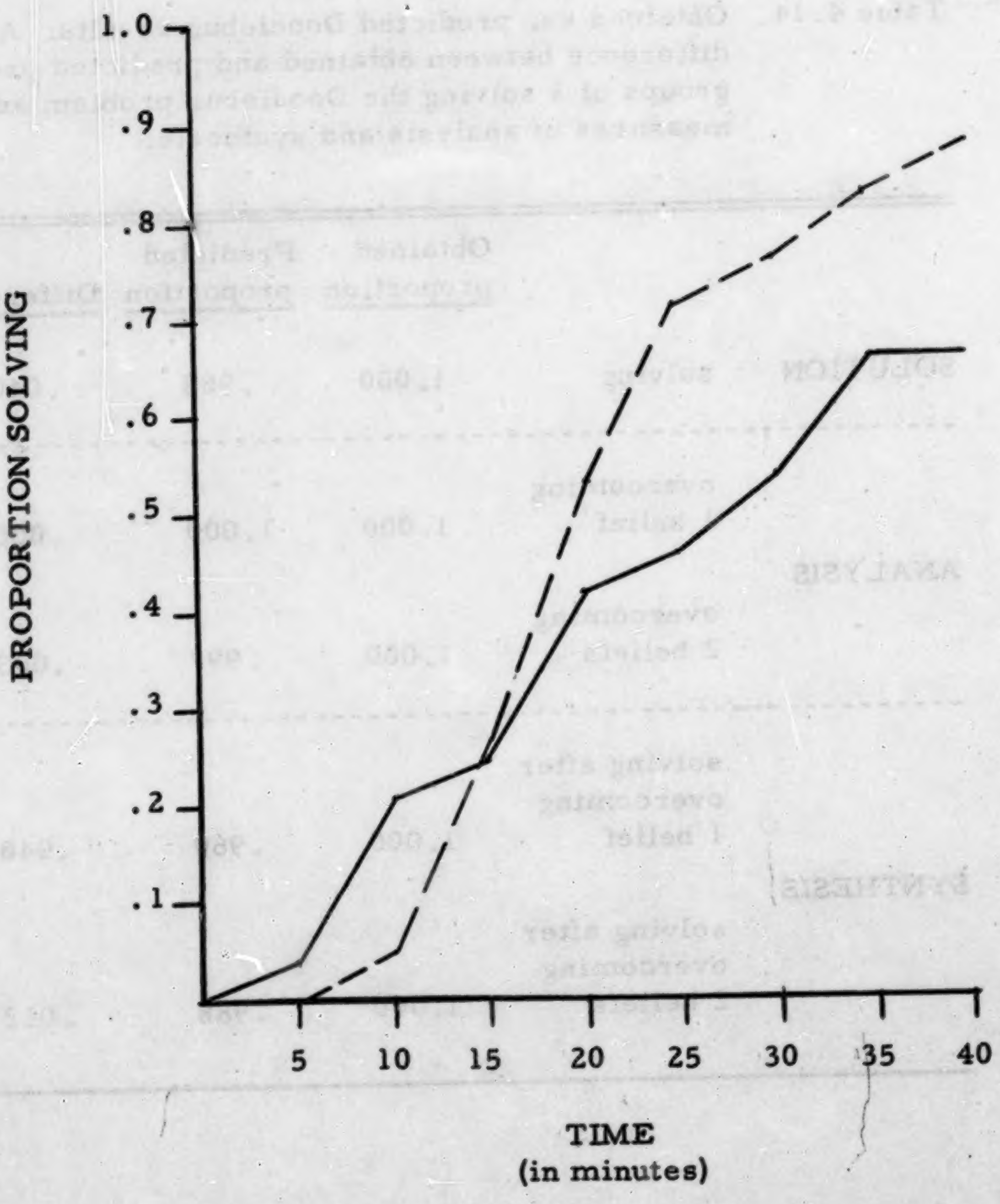

Figure 4.9. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving.

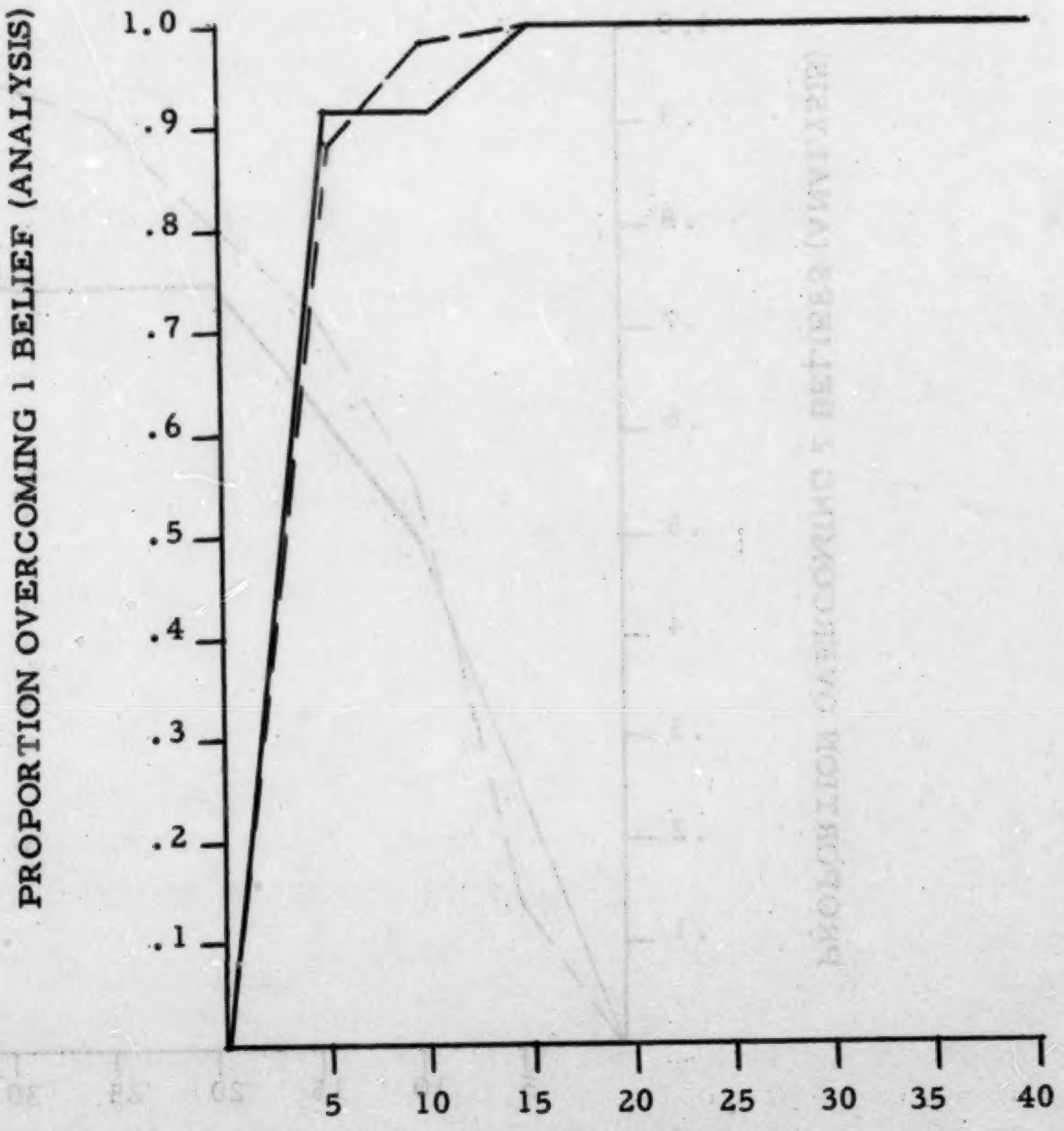

Figure 4.10. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 overcoming 1 belief.

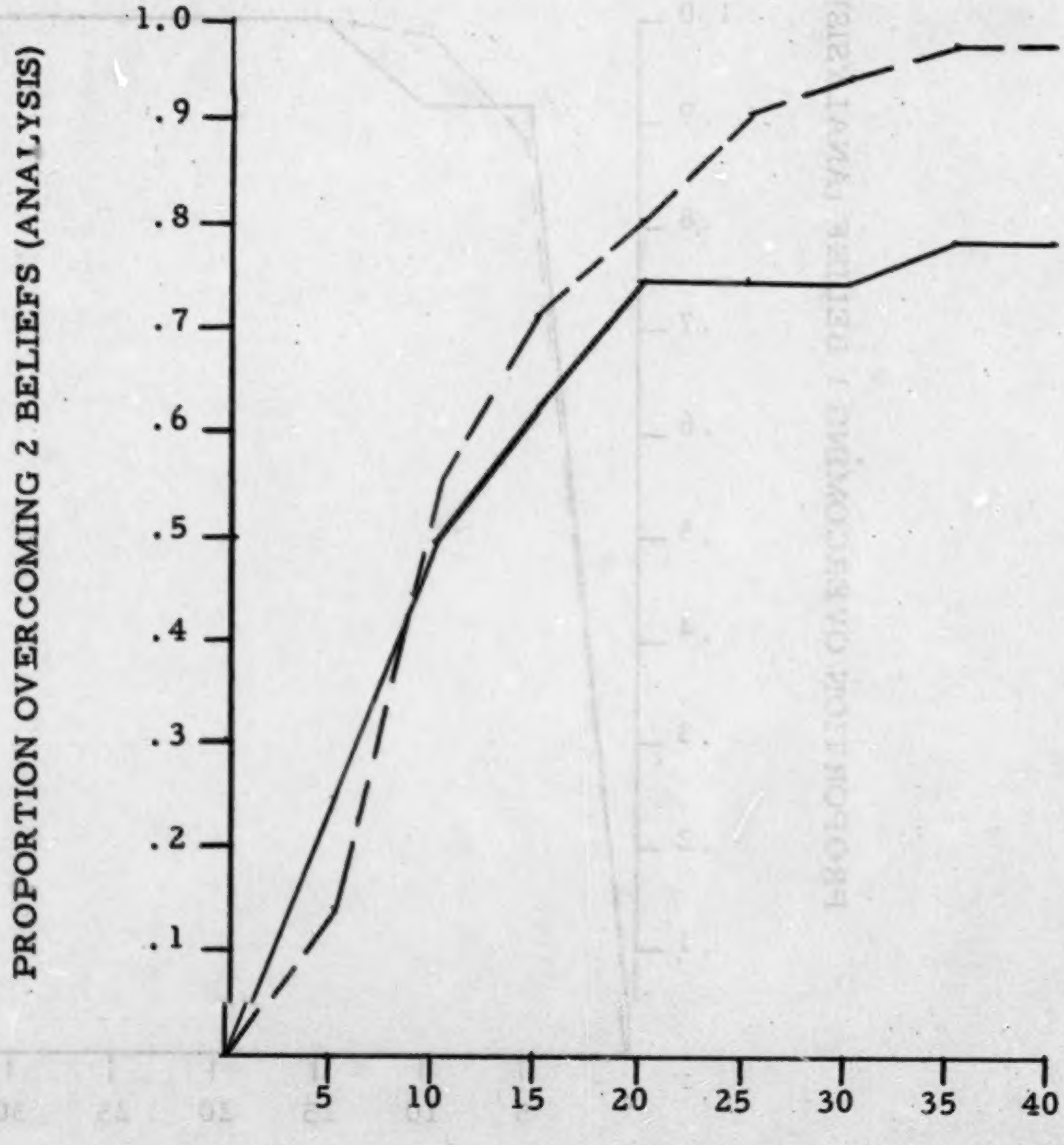

Figure 4.11. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 overcoming 2 beliefs.

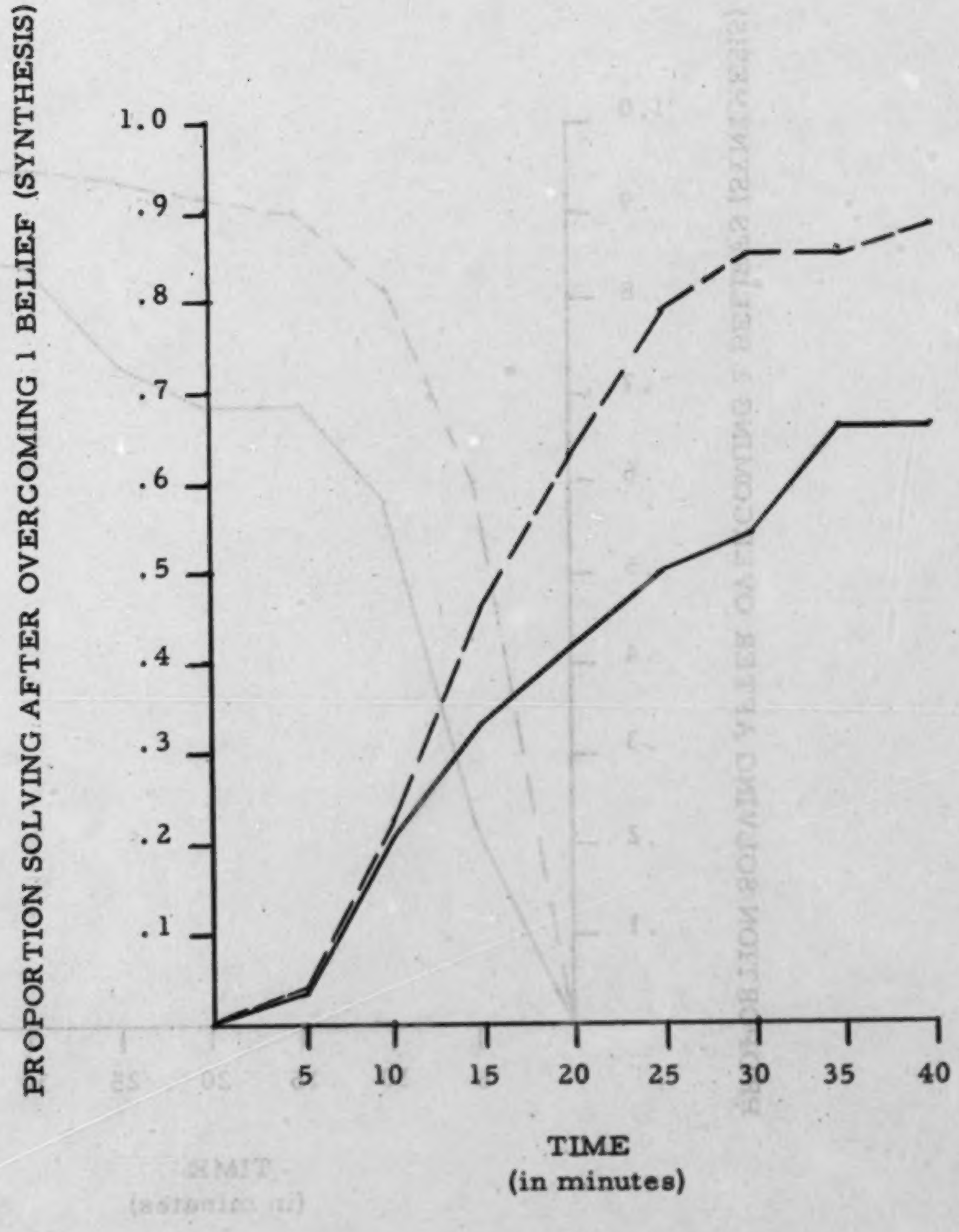

Figure 4.12. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving after overcoming 1 belief.

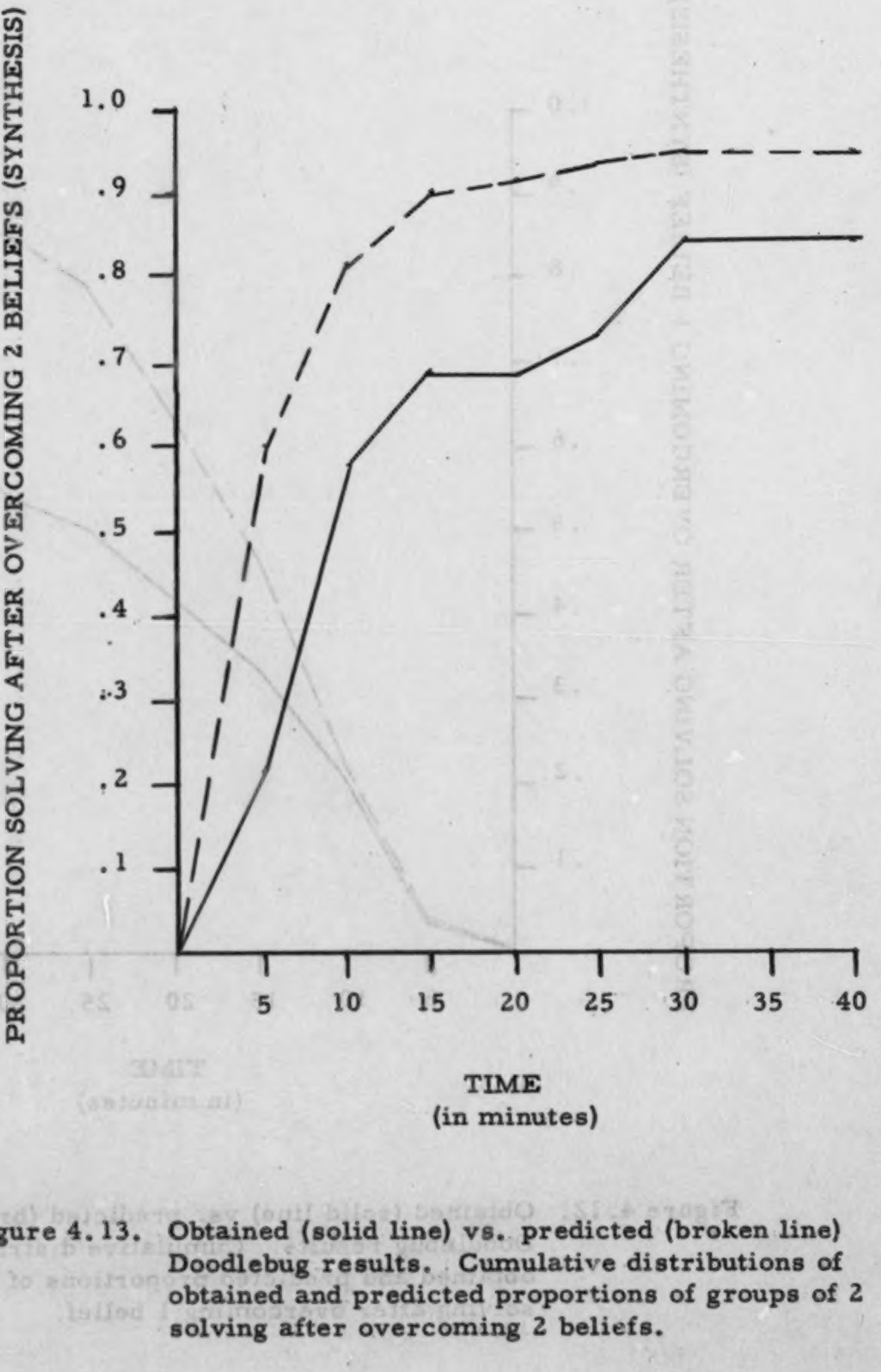

Figure 4.13. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 2 solving after overcoming 2 beliefs.

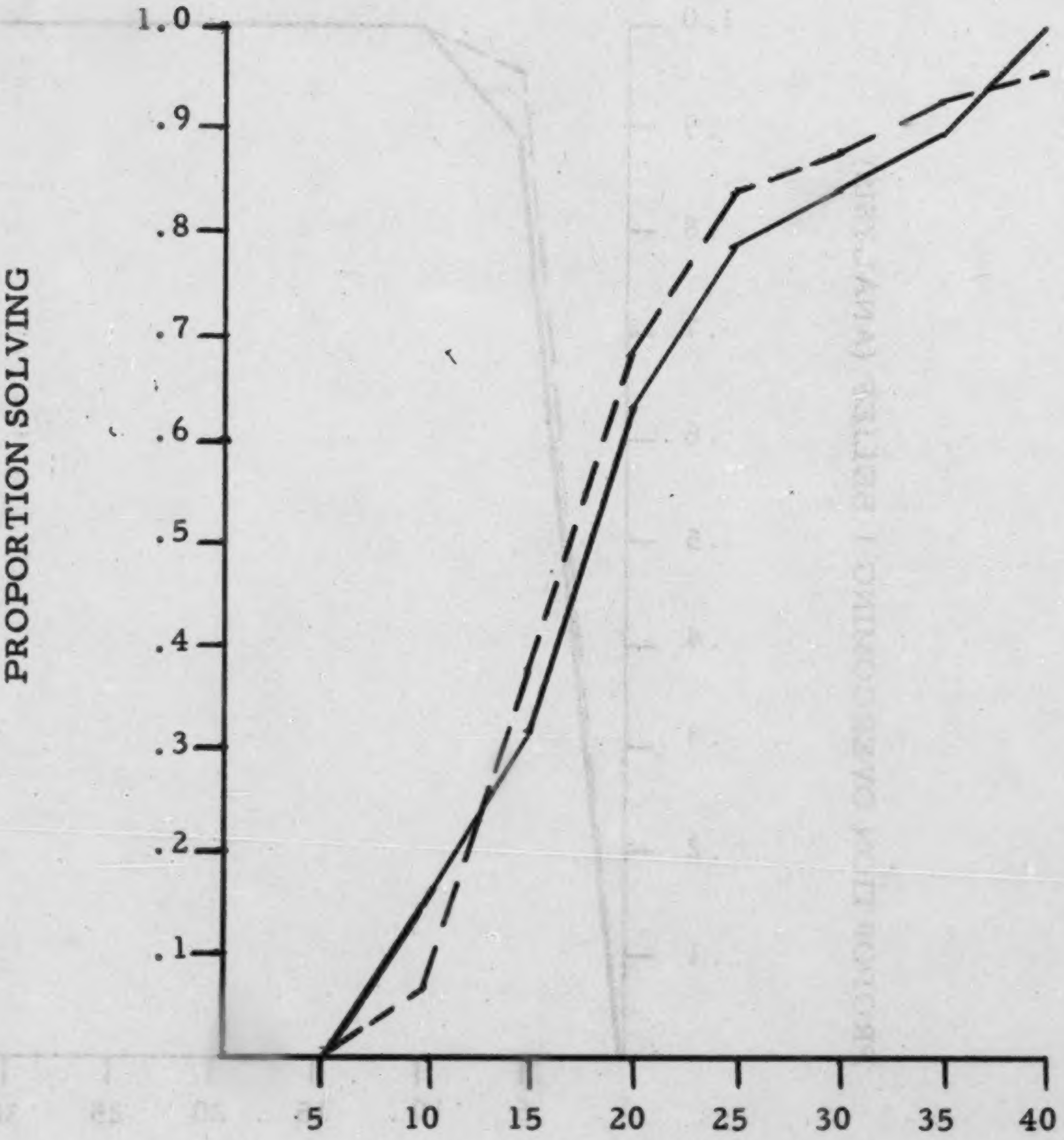

Figure 4.14. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained vs. predicted proportions of groups of 3 solving.

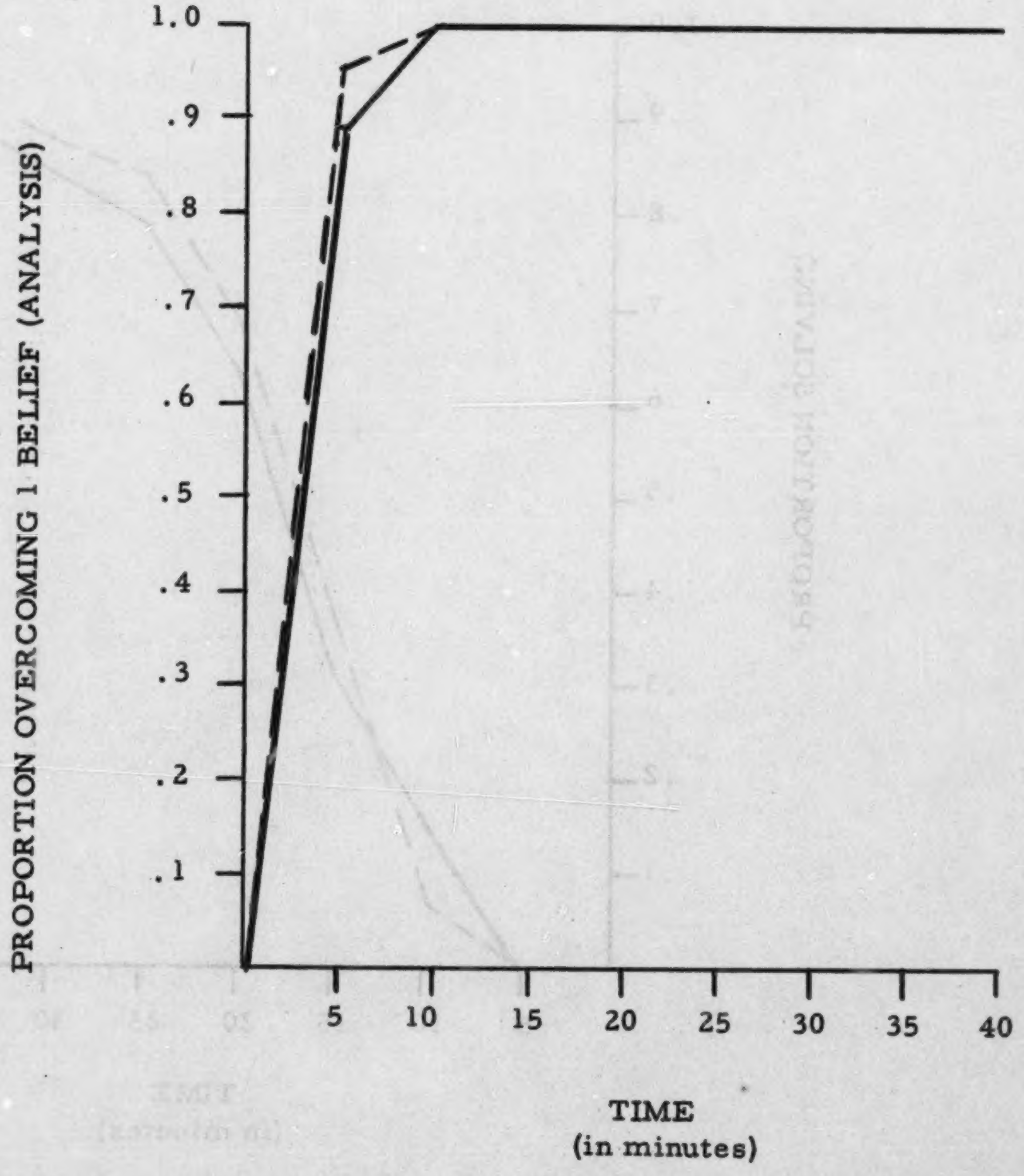

Figure 4.15. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 overcoming 1 belief.

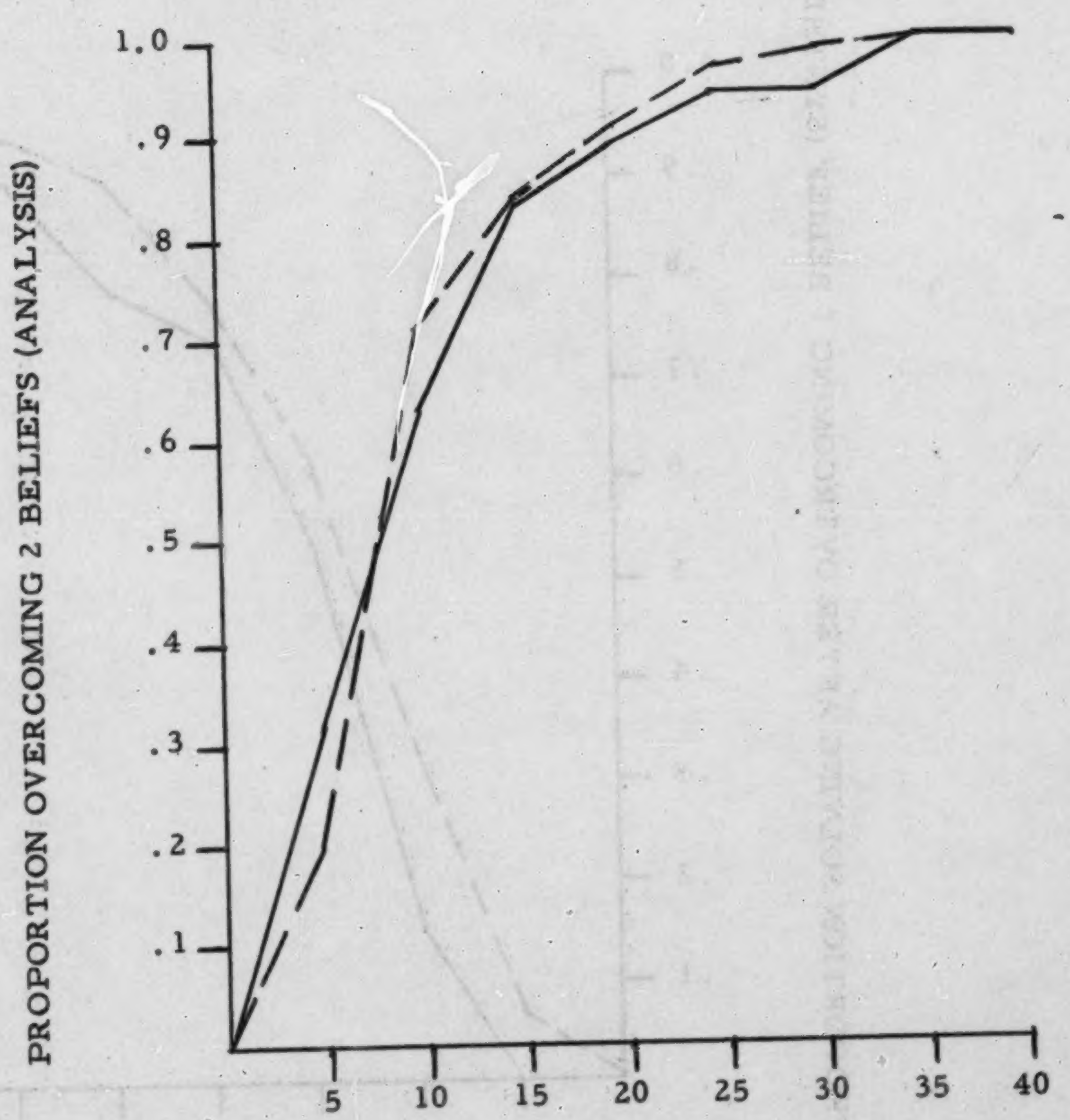

Figure 4.16. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 overcoming 2 beliefs.

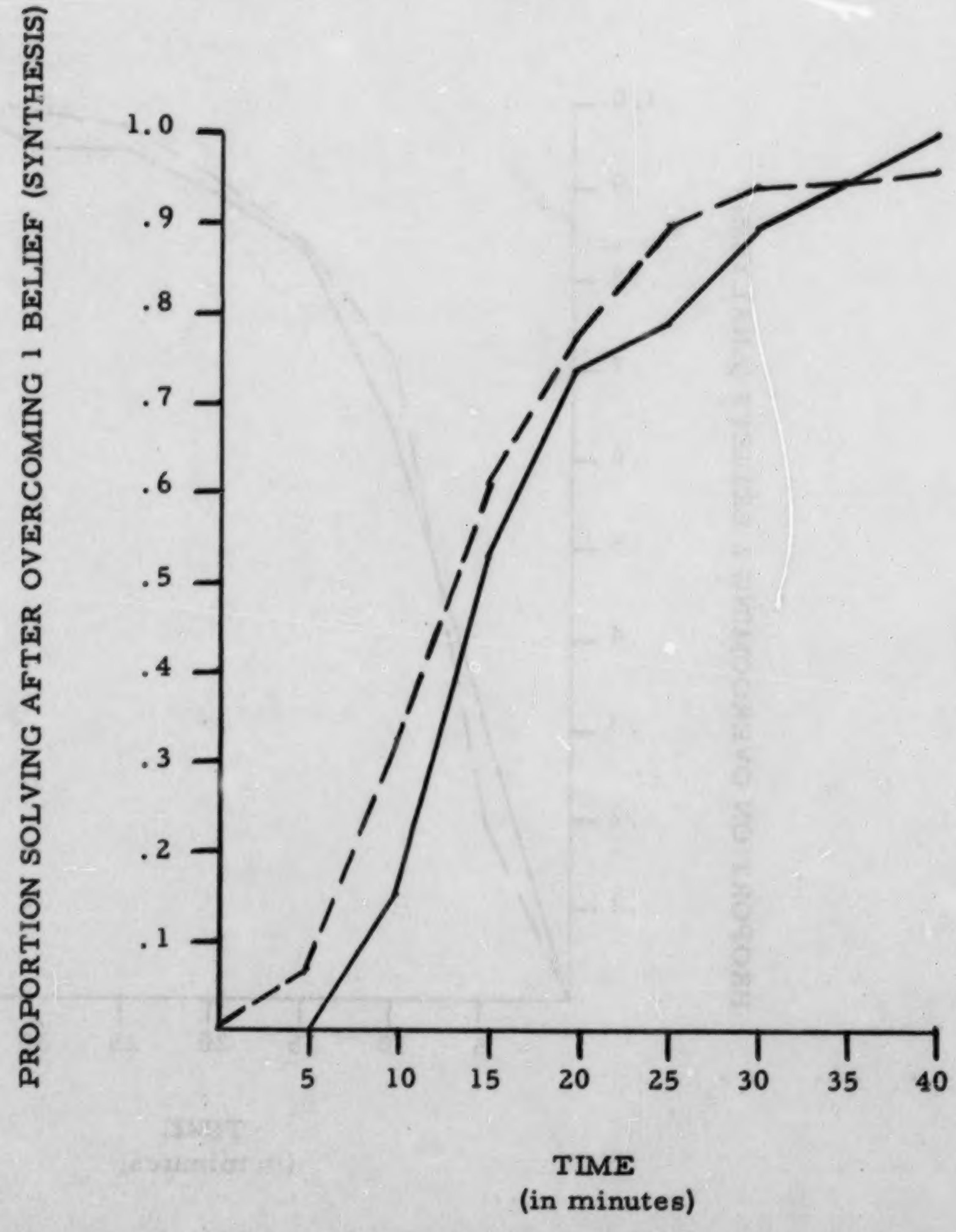

Figure 4.17. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 solving after overcoming 1 belief.

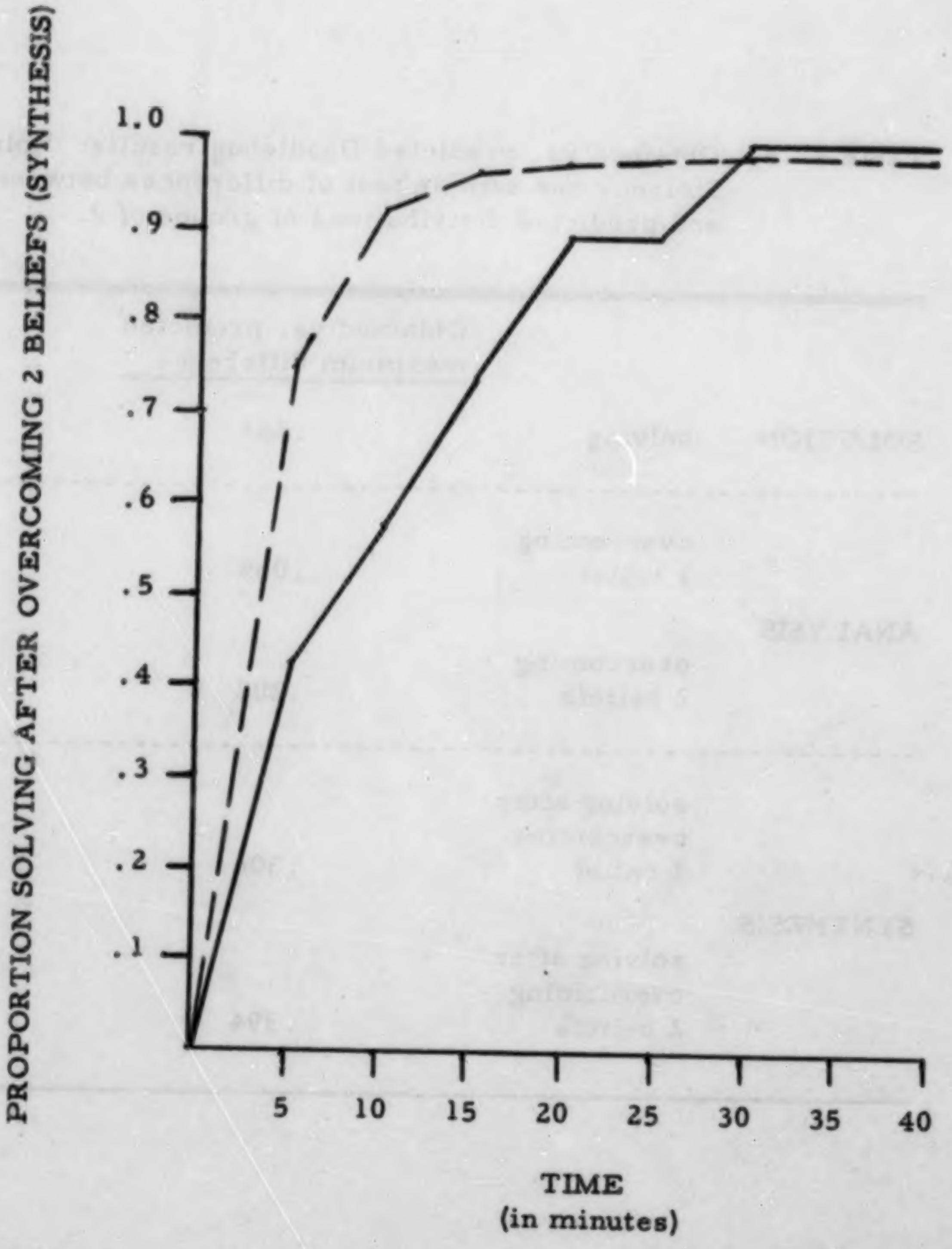

Figure 4.18. Obtained (solid line) vs. predicted (broken line) Doodlebug results. Cumulative distributions of obtained and predicted proportions of groups of 3 solving after overcoming 2 beliefs.

Table 4.15. Obtained vs. predicted Doodlebug results: Kolmogrov-Smirnov one sample test of differences between obtained and predicted distributions of groups of 2.

		Obtained vs. predicted maximum difference	p
SOLUTION	solving	.254	.05
ANALYSIS	overcoming 1 belief	.068	.20
	overcoming 2 beliefs	.202	.20
SYNTHESIS	solving after overcoming 1 belief	.306	.01
	solving after overcoming 2 beliefs	.394	.01

Not all the differences seen in the figures are significant at less than the .05 level, but the 2 measures of synthesis yield significant differences and the difference in proportions solving approaches the .05 level. The differences were in the direction indicating that the groups of 2 performed more poorly than predicted by the pooling model.

Table 4.16 presents the results of the Kolmogrov-Smirnov one sample test of the differences between the obtained and predicted results of groups of 3. The cumulative distributions of these measures are presented visually in figures 4.14 through 4.18. From the table it can be seen that on one measure of synthesis the obtained differed significantly at less than the .05 level from the predicted. All other measures yielded no differences at less than the .20 level of significance.

The obtained and predicted results may be summarized as follows. Generally, groups of 2 did not differ significantly from individuals, and the predicted performance of groups of 2 was higher than the actual performance. Groups of 3 generally performed superior to individuals and groups of 2, and in most cases their performance was as the pooling model predicted. The one exception was in the cumulative predictions of the second measure of synthesis. This comparison is more relevant than the comparison of proportions at the end of the 40 minute time limit because it takes into account the proportions at each of the 5 minute intervals. Therefore, we might expect that the lack of a significant difference between the obtained and predicted synthesis measure at the end of 40 minutes is spurious. By examining figure 4.18 it can be seen that the theoretical distribution reaches an asymptote much sooner than does the obtained distribution. This suggests that the lack of a significant difference between the obtained and predicted synthesis measure at the end of 40 minutes occurs because the comparison is made "too late." The lower performance of groups of 3 on the second measure of synthesis does suggest that in groups of 3 there is some inhibition of the synthesis process.

Table 4.16. **Obtained and predicted Doodlebug results: Kolmogrov-Smirnov one sample test of differences between obtained and predicted distributions of groups of 3.**

		Obtained vs. predicted maximum difference	**P**
SOLUTION	**solving**	**.065**	**.20**
ANALYSIS	**overcoming 1 belief**	**.112**	**.20**
	overcoming 2 beliefs	**.088**	**.20**
SYNTHESIS	**solving after overcoming 1 belief**	**.165**	**.20**
	solving after overcoming 2 beliefs	**.342**	**.05**

However, this inhibition is not enough to seriously effect the solution of the problem because the model predicts quite well the proportions of solutions.

The most troublesome result was the performance of groups of 2 on the Doodlebug problem. For most purposes they performed just like individuals--neither better nor worse. For the Lorge and Solomon model to predict accurately the best solver in the group must work on the problem without interference (or facilitation) from the other group members. If all the group members work simultaneously on the problem, this guarantees that the best as well as the worse solver is not meeting interference. Also, if the poorer solvers of the group simply follow the best solver throughout the problem solving process, the best solver will not be interfered with and the Lorge and Solomon model should give an accurate prediction. If all the group members are not working on the problem simultaneously and the one or more persons who are dominating the process are not the best solvers in the group, the best solver's behavior will be inhibited. This will lead to an over prediction by the Lorge and Solomon model.

Protocols of the group problem solving process were recorded by the experimenter and it was possible to examine whether or not one person was dominating the problem solving process or whether each of the group members were participating about equally. These protocols recorded who in the group overcame the beliefs and who solved the problem. Therefore, it was possible to examine whether or not one participated more in the solution of the problem than would be expected by chance if both members were participating equally. This was accomplished in the following way. In the protocols of the Doodlebug solutions there were 3 measures of participation, overcoming belief 1, overcoming belief 2, and solving the problem. In groups of 2 if both persons were participating equally, the probability of one person overcoming belief 1 would be .50. Similarly, the probability of one person overcoming belief 2 would be .50 and solving the problem would be .50. Therefore, the probability of one particular person doing all three things would be (.50)(.50)(.50) = .125 since the probabilities are independent. The probability of either person doing all 3 things would be .125 + .125 = .25. From the total number of solutions it would be expected by chance that in .25 of them one person would contribute all 3 parts if both persons in the groups were participating equally. The obtained proportion of groups in which one person did all 3 things was .50 (8 of the 16 groups who solved). A test of the difference between the obtained and predicted proportions (Walker and Lev, 1953) demonstrated that it was significant at less than the .02 level (z = 2.358). This indicates that the group members were not participating equally, and since the Lorge and Solomon model over predicts the performance of groups of 2, it suggests that the person who dominates the problem solving is not always the best solver. Apparently, dominance in the problem solving process of groups of 2 is determined by some personality variable which is not related to the skills most appropriate for the solution of the Doodlebug problem.

To examine participation in groups of 3 similar obtained and predicted proportions were compared. In groups of 3 if each person was participating equally, the probability of 1 person overcoming belief 1 would be .33 and, similarly, overcoming belief 2 would be .33 and solving the problem would be .33. Therefore, the probability of one person doing all three things would be (.33)(.33)(.33) = .0359, and the probability of any of the 3 persons doing all 3 things would be (.0359)(3) = .108. Therefore, of the total 19 solutions to the problem, it would be expected by chance that in .108 of them one person would contribute all 3 parts. The obtained proportion was .053. By comparing the obtained to the predicted proportion of .108, it can be seen that the obtained proportion is somewhat smaller than would be expected by chance indicating that the groups of 3 did not tend to have one person dominant. (A test of the difference between the proportions yielded a z less than 1.) The tendency of equal participation in groups of three suggests that the group members were working simultaneously on the problem which enabled the best solver to reach a solution unimpeded.

The above results indicate an odd difference between groups of 2 and groups of 3. In groups of 2 one person tends to dominate and that person is not always the person with the best problem solving skills, and in groups of 3 one person does not tend to dominate suggesting simultaneous problem solving by all the group members.

CONCLUSIONS

In the recall experiment the first thing which was examined was the variances of groups and individuals. It was found that the overall variance of groups of 3 was less than groups of 2 (although not quite significant) and both these variances were less than that of individuals. However, we cannot conclude that the differences in variances are due solely to the difference in the number of people working on the task. The groups performed near the ceiling of 8 correct words on the later trials and, therefore, their variances were less than individuals who tended not to reach the ceiling.

The hypothesis of group superiority and size received some support in the recall experiment. Groups of 2 and 3 were superior to individuals on the recall task. Furthermore, groups of 3 tended to perform superior to groups of 2 although this difference was not quite significant at the .05 level. From rather limited data on group size, that is, only groups of size 2 and 3 were examined, there appears to be a direct relationship between the size of the group and performance since groups of 2 were superior to individuals and groups of 3 tended to be superior to both individuals and groups of 2. Superiority increases with size (see figure 4.1).

The predictions of group performance based on the pooling model were accurate in most instances. Four of 20 predictions differed significantly from the obtained results. However, when the serial position effect which occurred on the first trial was taken into account in making the predictions, only 2 of the 20 predictions were faulty. Both of these occurred in the predictions of groups of 3. In both the first and last trial groups of 3 performed more poorly than that which was predicted by the pooling model (see figure 4.3). Examining figure 4.3 it can be seen that the overall performance by groups of 3 tends to be slightly below that predicted. This suggests that there is some slight interference occurring in groups of this size on the recall task. While there is no overall difference as indicated by the t-test of the difference of total mean correct, there might be some confounding resulting from having several trials when the groups were performing nearly perfectly. It is possible that there is a serious discrepancy between the predicted and obtained which only evidences itself on the first and last trials. When the serial position effect is taken into account for groups of 2, on no trial does the obtained differ significantly from the predicted. An alternative interpretation of the two significant differences in groups of 3 is that these differences occurred by chance. If it were given that there were no differences between the obtained and predicted results, it would be expected by chance that 1 of the 20 predictions would be significant at the .05 level. Therefore, it is not supposing a very odd occurrence to suggest that the 2 of 20 differences were due to chance. Which of these alternative explanations is correct cannot be determined by the present research.

A further problem concerning the adequacy of the prediction of groups of 3 on the recall task is that groups of 2 and 3 did not quite differ significantly ($p = .07$). This suggests that using the performance of groups of 2 to predict groups of 3 or vice versa cannot be rejected by the t-test. This further suggests that there is some kind of interference occurring in groups of 3 and they are performing somewhat like groups of 2.

While there is a suggestion of interference in groups of 3, nowhere is there an indication of facilitation, and it must be said that one need not resort to an explanation of group superiority which is based on the facilitative effects of interaction.

Examining figures 4.2 and 4.3 it can be seen that there is a greater tendency for the model to over predict groups of 3 than groups of 2. Perhaps as the group size increases there is more and more interference. If this is the case, adding a constant to the Lorge and Solomon model would be necessary for it to give an accurate prediction of group performance. The formula would then be: $P_g = 1 - (1 - P_i)^{\theta k}$, where θ is a constant less than 1 representing interference. To determine the numerical value groups of a greater size then 3 would have to be examined and the constant estimated from their performance.

As with the recall experiment, in the Doodlebug experiment there is also no evidence for facilitation, but there are further complications. The apparent direct relationship between group size and superiority is not found in the Doodlebug experiment as it was in the recall experiment. Groups of 3 were significantly superior to individuals, but this was not true of groups of 2. Groups of 2 were slightly superior to individuals in solutions to the Doodlebug problem. However, this difference was quite small and did not approach significance. On the other hand, groups of 3 were significantly superior to individuals and groups of 2 in proportions of solutions (see tables 4.10 through 4.12).

Of the 2 measures of analysis of the Doodlebug problem only one appears fruitful, that of the proportion overcoming two beliefs because the other measure, that of the proportion overcoming one belief was 1.00 in the individual, group of 2 and group of 3 conditions, the ceiling was reached in all 3 conditions. The differences in the proportions overcoming 2 beliefs indicates the same superiority that was found with the proportion solving the problem (again see tables 4.10 through 4.12). Groups of 2 were not significantly superior to individuals on this measure of analysis, whereas, groups of 3 were significantly superior to both individuals and groups of 2.

Examining the measures of synthesis as presented in tables 4.10 through 4.12 it may be seen that the first measure, the proportion who solve after overcoming 1 belief, is the same as the proportion who solve the problem since all the groups and individuals overcame one belief. The second measure of synthesis indicates that groups of 2 do not differ significantly from individuals but groups of 3 do differ significantly from individuals. The difference between groups of 3 and 2 is not quite significant ($p = .07$).

Generally, the measures of analysis, synthesis and solution indicate that groups of 2 are not significantly superior to individuals, whereas, groups of 3 do differ significantly from individuals and in most cases also from groups of 2. Therefore, the conclusion that groups of 3 are superior to individuals and groups of 2 can be made.

Comparing the obtained Doodlebug results with the results predicted using the pooling model indicates that groups of 2 perform significantly more poorly than predicted (see tables 4.13 and 4.15). Table 4.13 indicates that groups of 2 do significantly more poorly on analysis, but not significantly more poorly on the second measure of synthesis. On the other hand, table 4.15 indicates the converse (poorer synthesis but not analysis). The most reasonable conclusion is that the performance of groups of 2 in both analysis and synthesis is inferior to that which is predicted by the pooling model. Also from the significant differences between obtained and predicted proportions of solutions, we can conclude that groups of 2 perform inferior to that which the pooling model predicts (see tables 4.13 and 4.15).

In most of the comparisons between obtained and predicted, groups of 3 perform as the pooling model predicts (see tables 4.14 and 4.16). The arcsin test of the differences between the obtained and predicted proportions at the end of 40 minutes does not yield any significant differences. However, when the Kolmogrov-Smirnov test tests the difference between predicted and obtained cumulative distributions, the obtained second synthesis measure is significantly inferior to the predicted. The Kolmogrov-Smirnov test is the better indicator of what is occurring since it takes into account the total distribution, whereas, the arcsin test takes into account only the performance at the end of 40 minutes, and by that time the obtained proportion had reached the limit of 1.00. It can be concluded that the pooling model is an adequate explanation for analysis by groups of 3. For synthesis, it may be concluded that there is interference occurring in groups of 3 such that they do not perform as well as predicted, but this interference is not strong enough to produce a significant difference between the obtained and predicted proportion of solutions by groups of 3. Groups of 3 were superior to individuals, but an explanation of this superiority which is based on facilitative effects of group interaction need not be resorted to since the pooling model can account for the superiority, and if anything, there was interference occurring in groups of 3. Groups of 2, however, are not significantly superior to individuals and do not perform as well as predicted by the pooling model which indicates that something is occurring in the groups of 2 which interferes in some serious way with pooling.

An examination of the protocols of the problem solving by groups of 2 demonstrated that one member of the group dominated the problem solving. Furthermore, the Lorge and Solomon model over predicted their performance. This indicates that the dominant person was not always the best problem solver. It is likely that in many of the groups of 2 the best problem solver had his behavior inhibited by the other group member. Also, this suggests that dominance in this situation is unrelated to problem solving ability. In groups of 3 dominance by one member did not occur indicating that the group members were working simultaneously on the problem since their performance was accurately predicted in most instances.

It is difficult to suggest how the Lorge and Solomon model might be revised to better predict behavior of groups on the Doodlebug problem. In a sense, no model is needed to predict how groups of 2 perform since they behave like individuals. The application of the model, however, permits the discovery of interference. Whether this interference is particular to groups of 2, or groups containing an even number of members, or the Doodlebug problem cannot be determined by this study. Taylor and Faust's (1952) experiment found a lack of superiority of groups of 2 over individuals, whereas, groups of 4 were superior. This suggests that even size groups do not behave as individuals, but whether or not there is more interference in groups of even size than groups of odd size is impossible to say. Further research with groups of varying size and different problems must be done before definite conclusions of group size and interference can be made. Also, further research may lead to ways in which the Lorge and Solomon model can be revised to be better able to predict group performance.

We may conclude that the pooling of abilities model does not predict all of the group behavior studied by these experiments. On a simple task of memorizing nonsense words it appears that group interaction neither facilitates nor seriously disrupts performance in groups of 2 and 3, although there is a slight indication of interference in groups of 3. On a complex task of solving the Doodlebug problem, the pooling model over predicts the behavior of groups of 2. Furthermore, groups of 2 do not differ significantly from individuals. For groups of 3 and model predicts quite well except for synthesis where interference does occur. We can conclude that there is not a simple, direct relationship between group size and performance on this task and also in groups of 2, one person dominance occurs which is unrelated to problem solving ability resulting in interference. This does not occur in groups of 3 and apparently there is much less interference occurring in groups of this size.

The idea that group interaction is facilitative of individual performance receives no support from either of the 2 experiments. However, there is evidence of interference. In experiment 1 the interference was slight. In experiment 2 there is strong evidence for interference in groups of 2, and in groups of 3 interference occurs in the synthesis part of the problem solving process. However, this interference is not enough to seriously disrupt the solution of the problem.

SUMMARY

Two experiments were performed testing a mathematical model of group performance. This model, devised by Lorge and Solomon (1955), says that the probability of a group solving a problem is greater than the probability of an individual solving a problem simply because there is more than one individual in a group. Therefore, group superiority does not have to be a result of any facilitative qualities of the group situation. Furthermore, the model allows for a prediction of group behavior or a proper baseline against which to judge whether or not there is facilitation or interference.

The first experiment tested the model using as the task the recall of nonsense words. Groups of 2, groups of 3 and individuals were given 9 recall trials on 8 nonsense words. In this experiment to apply the Lorge and Solomon model each nonsense word was treated as a problem to solve. The results of this experiment demonstrated that groups of 2 and groups of 3 were superior to individuals. Groups of 3 were not quite significantly superior to groups of 2. Also, in all conditions there was a strong serial position effect. When this serial position effect was taken into account the predictions of groups of 2 and 3 were quite accurate. However, the model slightly over predicted the performance of groups of 3 which may indicate a tendency for interference in groups of this size. It was concluded that an explanation of group superiority which uses the facilitative effects of the group is not suitable since the superiority can be explained more simply by the Lorge and Solomon model.

In the second experiment groups of 2, groups of 3 and individuals were given a complex problem (Doodlebug problem) to solve.

The results showed groups of 3 to be superior to individuals and groups of 2. Groups of 2 were not significantly different from individuals. The Lorge and Solomon model considerably over predicted the performance of groups of 2 in analysis, synthesis and solution to the Doodlebug problem. This indicates the presence of interference in groups of this size. The predictions of the performance of groups of 3 was accurate except for the synthesis phase of problem solving. In this phase groups of 3 performed inferior to that predicted by the Lorge and Solomon model suggesting interference. However, the interference was not strong enough to disrupt the prediction of the solution.

The strong interference present in groups of 2 seemed to be due to one person, regardless of his problem solving ability, dominating the solution to the problem. This was supported by comparing expected and obtained participation values.

Again, in the second experiment there was no indication of group facilitation. Group superiority, where it occurred, could be explained by the Lorge and Solomon model.

In both experiments there was some indication of interference in group performance. In the first experiment the interference was slight. In the second experiment the interference was strongest in groups of 2 and present somewhat in groups of 3. No evidence was found for group facilitation.

BIBLIOGRAPHY

Allport, F. H. Social psychology. Cambridge, Mass.: Riverside, 1924.

Allport, F. H. A structuronomic conception of behavior: Individual and collective I. Structural theory and the master problem of social psychology. J. abnorm. soc. Psychol., 1962, 64, 3-30.

Anderson, N. H. Group performance in an anagram task. J. soc. Psychol., 1961, 55, 67-75.

Asch, S. E. Social psychology. New York: Prentice-Hall, 1952.

Back, K. W. Influence through social communication. J. abnorm. soc. Psychol., 1951, 46, 9-23.

Bales, R. F. Interaction process analysis: A method for the study of small groups. Cambridge, Mass.: Addison-Wesley, 1950.

Barnlund, D. C. A comparative study of individual, majority and group judgment. J. abnorm. soc. Psychol., 1959, 58, 55-60.

Coch, L. and French, J. R. P., Jr. Overcoming resistance to change. Hum. Relat., 1948, 1, 512-532.

Davis, J. H. Models for the classification of problems and the prediction of group problem-solving from individual results. Unpublished Ph. D. dissertation, Michigan State University, 1961.

Dashiell, J. F. Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (Ed.), Handbook of social psychology. Worcester: Clark Univ. Press, 1935.

Duncan, C. P. Recent research on human problem solving. Psychol. Bull., 1959, 56, 397-429.

Edwards, A. L. Statistical methods for the behavioral sciences. New York: Rinehart, 1956.

Ekman, G. The four effects of co-operation. J. soc. Psychol., 1955, 41, 149-162.

Faust, W. L. Group vs. individual problem solving. J. abnorm. soc. Psychol., 1959, 59, 68-72.

Gordon, Kate. Group judgments in the field of lifted weights. J. exp. Psychol., 1924, 7, 389-400.

Gurnee, H. A comparison of collective and individual judgments of fact. J. exp. Psychol., 1937a, 21, 106-112.

Gurnee, H. Maze learning in the collective situation. J. Psychol., 1937b, 3, 437-444.

Gurnee, H. The effect of collective learning upon the individual participants. J. abnorm. soc. Psychol., 1939, 34, 529-532.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951.

Husband, R. W. Cooperative versus solitary problem solution. J. soc. Psychol., 1940, 11, 405-409.

Jenness, A. The role of discussion in changing opinion regarding a matter of fact. J. abnorm. soc. Psychol., 1932, 27, 279-296.

Johnson, D. M. The psychology of thought and judgment. New York: Harpers, 1955.

Kelly, H. H. and Thibaut, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), Handbook of social psychology. Cambridge, Mass.: Addison-Wesley, 1954, 734-785.

Lewin, K. Group decisions and social change. In Eleanor E. Maccoby, T. M. Newcomb, and E. L. Hartley (Eds.), Readings in social psychology. New York: Holt, 1958.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Cambridge, Mass.: Riverside, 1953.

Lorge, I. and Solomon, H. Two models of group behavior in the solution of Eureka-type problems. Psychometrika, 1955, 20, 139-148.

Lorge, I., Fox, D., Davitz, J. and Brenner, M. A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. Psychol. Bull., 1958, 337-372.

Marquart, Dorothy I. Group Problem solving. J. soc. Psychol., 1955, 41, 102-113.

Marr, J. N. Unpublished research. Michigan State University. 1961.

Meumann, E. Haus- und Schularbeit. Leipzig, Klinkhardt, 1914. Reported in Allport, F. H. Social psychology. Cambridge, Mass.: Riverside, 1924.

Perlmutter, H. V. Group memory of meaningful material. J. Psychol., 1953, 35, 361-370.

Perlmutter, H. V. and DeMontmollin, Germaine. Group learning of nonsense syllables. J. abnorm. soc. Psychol., 1952, 47, 762-769.

Restle, F. Speed and accuracy of cognitive achievement in small groups. Symposium, Social Science Research Council, 1961.

Restle, F., and Davis J. H. Success and speed of problem solving by individuals and groups. (In press, Psychol. Rev., 1962)

Rokeach, M. The open and closed mind. New York: Basic Books, 1960.

Shaw, Marjorie E. Comparison of individuals and small groups in the rational solution of complex problems. Amer. J. Psychol., 1932, 44, 491-504.

Sherif, M. Group influences upon the formation of norms and attitudes. In Eleanor E. Maccoby, T. M. Newcomb and E. L. Hartley (Eds.), Readings in social psychology. New York: Holt, 1958.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Stroop, J. B. Is the judgment of the group better than that of the average member of the group? J. exp. Psychol., 1932, 16, 550-560.

Taylor, D. W., Berry, P. C., and Block, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? Admin. science Quart., 1958, 3, 23-47.

Taylor, D. W., and Faust, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. J. exp. Psychol., 1952, 44, 360-368.

Thibaut, J. W., and Kelly, H. H. The social psychology of groups. New York: Wiley, 1959.

Thorndike, R. L. The effects of discussion upon the correctness of group decision when the factor of majority influence is allowed for. J. soc. Psychol., 1938, 9, 343-362.

Timmons, W. M. Can the product superiority of discussors be attributed to averaging and majority influences? J. soc. Psychol., 1942, 15, 23-32.

Walker, Helen M. and Lev, J. Statistical inference. New York: Holt, 1953.

Watson, G. B. Do groups think more efficiently than individuals? J. abnorm. soc. Psychol., 1928, 23, 328-336.

Zajonc, R. B. A note on group judgments and group size. Hum. Relat., 1962a, 15, 177-180.

Zajonc, R. B. The effects of feedback and probability of group success on individual and group performance. Hum. Relat., 1962b, 15 149-161.

Reprinted from the
Journal of Abnormal and Social Psychology,
1962, Vol. 65, No. 1, 64-67

MEMORIZING BY INDIVIDUALS AND GROUPS:

A Test of the Pooling-of-Ability Model[1]

Ronald A. Hoppe[2]

Michigan State University

Perlmutter and DeMontmollin (1952) compared group learning of nonsense words (paralogs) with individual learning of the same nonsense words. Their procedure was to present a list of 19 two-syllable nonsense words and the task was to recall the words at the end of each presentation. Individuals and groups of three were presented with the same task. Learning curves were constructed for both the groups and the individuals. The groups started at a higher level of recall and remained at a higher level throughout the five learning trials. At the fifth trial the groups were averaging 18 of the 19 words, and the individuals were averaging 11 of the 19. The difference between groups and individuals was significant, and Perlmutter and DeMontmollin concluded that groups of three were superior to individuals in learning a list of nonsense syllables. Perlmutter and DeMontmollin invoked an elaborate explanation involving interaction of the group members in order to account for the superiority of the group over the individual. However, the general results might be explained on the basis that individuals give only the words they separately remember, whereas group performance reflects the pooled learning of the members, who can be thought of as remembering separately. Perlmutter and DeMontmollin, at the time of their study, did not have a model available to test this hypothesis.

Lorge and Solomon (1955) offered a mathematical model of group problem solving to explain the results of Shaw (1932) who found that more three-member groups solved eureka-type problems than did individuals. According to the Lorge-Solomon model the probability that at least one of three individuals in a group will arrive at a solution is greater than the probability of success for one individual working separately. Let p_i be the probability that an individual will solve the problem within a given time. Then $(1 - p_i)$ is the probability that an individual will not arrive at the solution. If k individuals are working on the problem together, but each still has probability p_i of solving and the probabilities are independent, then the probability that none of them solves is $(1 - p_i)^k$. Therefore, the probability of the group arriving at the solution, p_g, is equal to $(1 - p_i)^k$. This model assumes no interaction which facilitates remembering in groups, but merely represents the pooling of ability of the members of the group.

---

[1] The research was supported by the United States Office of Education, Grant 71-2141, Milton Rokeach, director. The author wishes to acknowledge the advice and criticism of Frank Restle who was a continuous help throughout the present research.

[2] Now at Miami University.

The notion of pooling of abilities might also apply to Perlmutter and DeMontmollin's (1952) results. While the group remembered more nonsense words on each trial than did the individual, the group results may have been a pooling of words remembered by each of the members, and not attributable to an further facilitation from the group.

The purpose of this study is to replicate Perlmutter and DeMontmollin's (1952) study and to test the hypothesis that the superiority of the group can be explained by the Lorge and Solomon (1955) pooling model.

To apply the Lorge and Solomon (1955) model to group recall, the following consideration must be made. Treat the probability of recalling a previously seen nonsense word within a given time just as Lorge and Solomon treat the probability of solving a problem within a given time. Let p_i = the probability that an individual will recall a given word. Then the probability of a group, p_g, of size k of recalling the word would equal $1 - (1 - p_i)^k$.

To predict Perlmutter and DeMontmollin's (1952) results directly from this model it would be necessary to know the individual and group results of recalling each of the words in their list, but data were not available. However, a prediction of their results is possible if it is assumed that each of the words in the list has an equal probability of being remembered. Making this assumption the proportion of the entire list recalled by the group, p_{tg}, can be estimated from the proportion of the entire list recalled by individuals, p_{ti}, by the following formula: $p_{tg} = 1 - (1 - p_{ti})^k$. It follows that the predicted group mean, $\bar{G}$, is found by multiplying p_{tg} by the number of words on the list.

Using the formula mentioned above, Perlmutter and DeMontmollin's (1952) group results were predicted. Only a part of Perlmutter and DeMontmollin's results was used since they had the following balanced design: two groups of subjects were tested both as groups and individuals. Group A was given five recall trials as individuals on a list of 19 nonsense words, and then five recall trials as groups on a different but comparable list. Group B was given a list to recall as a group followed by another to recall as individuals. We used the individual results of Group A to predict the group results of Group B. Table 1 describes the predicted and obtained means for each trial. It can be seen that the obtained and predicted results are quite close.

The present experiment was concerned with testing the Lorge and Solomon (1955) model in a group and individual recall setting similar to that of Perlmutter and DeMontmollin (1952). The present experiment differed from Perlmutter and DeMontmollin's in that (a) the number of words to be recalled was reduced from 19 to 8; (b) both individual and groups were given nine instead of five trials; (c) more individuals and groups were tested than by Perlmutter and DeMontmollin.

Method

Subjects

The subjects were 65 voluntary introductory psychology students: 41 females and 24 males, divided into 14 groups of 3, where each group contained only members of the same sex; and 23 subjects tested individually. Each group was tested separately and the individuals were tested either singly or in pairs.

Procedure

Both the groups and individuals were told that they would be presented with eight nonsense words to recall. They were told to attempt to learn the words as quickly as possible. Both were told to record the recalled words on a pad of paper in front of them, one word per page, and after they had turned the page they were not allowed to turn back. Each individual recorded his own words, and in the groups the subject who sat in the middle was asked by the experimenter to record his own words and the other members' recalled words. The groups were also told to discuss each recalled word.

Occasionally a group asked if they could divide the task with each member taking a certain proportion of the words. To this the experimenter replied, "No," and that this would give results of three individuals instead of group results, and the experiment was concerned with group results. The experimenter observed no apparent division of the task.

After the instructions were read, both individuals and groups were presented the stimuli in prearranged random orders.

The stimuli, constructed by combining three-letter nonsense syllables from Hilgard's (1951) list, were: BEMAW, HETIG, JALOB, KUQOM, LIPUF, MUWIY, SOHUJ, and ZADEC.

Groups and individuals were given nine trials with each of the above words presented once each trial. The words were presented in a different random order each trial and, furthermore, the order of the orders was randomized across individuals and groups. Each word was presented for 5 seconds followed by a 5-second interval in which no word was present. After the last word had been presented the subjects were given 2 minutes in which to recall the words. At the end of this period both the groups and individuals were told to mark an "X" on the next blank page, turn it, and await the next trial. At the end of the ninth trial the experimental session was concluded.

Results and Discussion

The group results were predicted from the individual results, using the Lorge and Solomon (1955) model in the same manner as Perlmutter and DeMontmollin's (1952) results were predicted. The assumption that each word had an equal chance of being recalled seemed tenable since the order of presentation of the words was randomized and the words were about equally meaningful. Figure 1 presents the mean number of words recalled correctly per trial compared with the predicted mean. An inspection of the figure suggests that the predicted and obtained results are quite close except for the first trial. The mean total number of words recalled by groups on all nine trials was 64.07 and the predicted mean total was 65.73. The difference of 1.66 was not significant ($t = 1.70$, $df = 13$, $p > .10$). The difference between the obtained and predicted results for each trial was also tested, using the t test. The only difference on the first trial ($t = 4.33$, $df = 13$, $p < .01$).[3]

---

[3] There is no adequate statistic available to test the closeness of the predictions. The t test was used but in so doing the variance of the predictions was assumed to be zero and the predicted mean was treated as a parameter. The predictions, since they were based on the individual results, are variable, but there was no adequate estimate of this variance. The assumption of zero variance of the predictions was overly stringent for the purpose of this study. While it is possible that the difference on the first trial would not be significant if the variance of the predictions was known, the difference is still large enough to be concerned with, and suggests a flaw in the theory.

The difference on the first trial indicates an apparent weakness in the assumption that all words are equally likely to be remembered. While the positions of the words were randomized on each trial, a serial position effect could still occur on the first trial. If so, the assumption that each word has an equal chance of being recalled is not tenable. Figure 2 shows the proportion of correct recalls by individuals at each of the eight serial positions on Trial 1. It can be seen that the first word presented was recalled by nearly 70% of the individuals while the words in the middle of the first presentation were recalled by less than 10% of the individuals, indicating a strong serial position effect.

What seemed to have happened on the first trial was that each of the group members tended to learn the same words. The pooling of abilities model with the assumption of each word having an equal chance of being recalled does not take this into account. The model pools the ability of all members with a disregard to which words were learned. An extreme example of this would be if all members of the groups and all individuals learned only the first and last words on the list. In this case the group would perform no better than the individual, but the prediction of the group's performance would be higher than the actual performance.

Because of the serial position effect, a new predicted mean, G, for the first trial was made, based on the proportion of individuals recalling the words in each of the eight positions. The formula, $p_g = 1 - (1 - p_i)^k$, was applied separately for the words of each position. The new prediction of the mean number of words recalled, $\bar{G}$, was made by summing the predicted proportions of the words recalled by the group in each position, $\bar{G} = \Sigma p_g$. The new prediction of 3.82 correct responses on Trial 1 was found not to differ significantly from the obtained value of 3.36 ($t = 1.49$, $df = 13$, $p > .10$).

From the above results it can be seen that the serial position effect can cause a discrepancy between the predicted and obtained group performance when the items in the list are lumped. The predictions of Perlmutter and DeMontmollin's (1952) results (Table 1) are somewhat high on early trials and low on later trials. If Perlmutter and DeMontmollin used the same order on each trial, the early high and later low predictions might be expected. They did not report whether or not thry used the same order on each trial. In the present experiment with a different order of presentation of words on each trial, most of the serial position effect is likely to be cancelled out after the first trial.

The results of the present experiment suggest that one need not introduce a concept of group facilitation or interaction in order to explain the group learning of nonsense words.

TABLE I

Obtained and Predicted Means of Perlmutter and DeMontmollin (1952)

Trial	Individual means	Obtained group means	Predicted group means
1	2.3	5.5	6.1
2	5.4	10.2	12.0
3	7.7	14.8	15.0
4	9.7	17.3	16.8
5	11.4	18.2	17.8

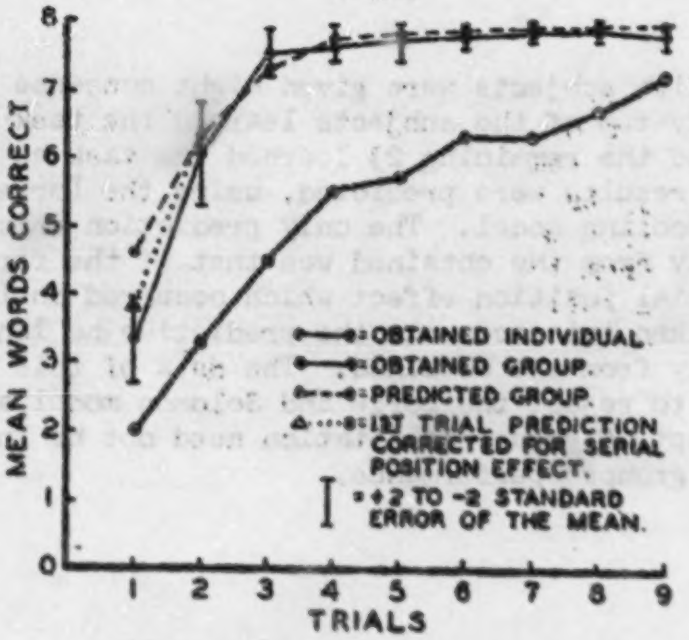

Fig. 1. Learning curves for individuals and groups, and group performance predicted from pooling of individual learning.

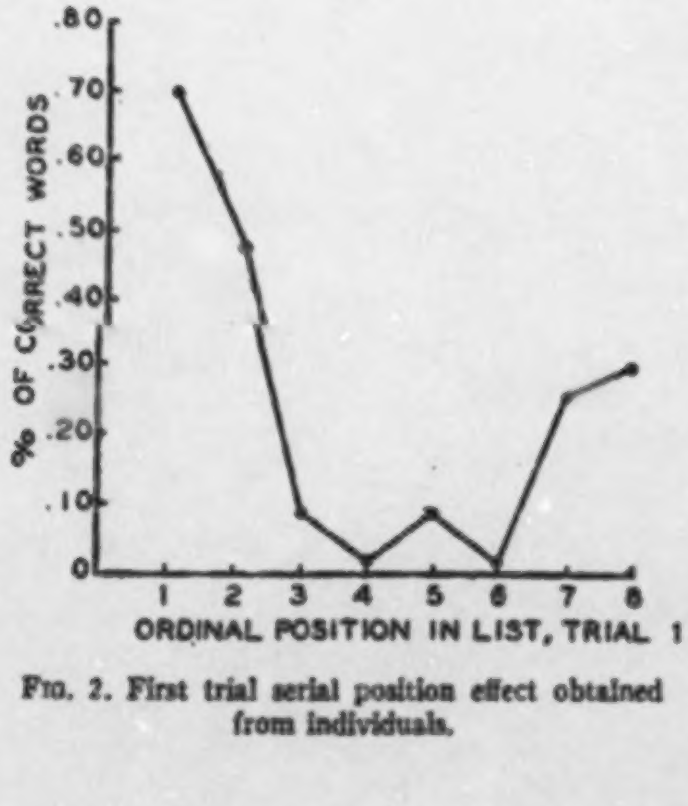

Fig. 2. First trial serial position effect obtained from individuals.

Summary

Sixty-five subjects were given eight nonsense words to learn. Forty-two of the subjects learned the task in groups of three, and the remaining 23 learned the task as individuals. The groups' results were predicted, using the Lorge and Solomon (1955) pooling model. The only prediction which differed significantly from the obtained was that of the first trial. When the serial position effect which occurred on the first trial was taken into account, the prediction no longer differed significantly from the obtained. The data of this study cannot be said to reject the Lorge and Solomon model and indicates that a concept of group facilitation need not be introduced to explain the groups' performance.

References

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951. Pp. 517-567.

Lorge, J., & Solomon, H. Two models of group behavior in the solution of eureka-type problems. Psychometrika, 1955, 20, 139-148.

Perlmutter, H. V., & DeMontmollin, G. Group learning of nonsense syllables. J. abnorm. soc. Psychol., 1952, 47, 762-769.

Shaw, Marjorie E. Comparison of individuals and small groups in the rational solution of complex problems. Amer. J. Psychol., 1932, 44, 491-504.

THE EFFECTIVENESS OF PROGRAMMED AND TRADITIONAL METHODS OF INSTRUCTION IN A COMPLEX VERBAL PROBLEM SOLVING SITUATION

By

Walter A. Watman

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1962

ABSTRACT

THE EFFECTIVENESS OF PROGRAMMED AND TRADITIONAL METHODS OF INSTRUCTION IN A COMPLEX VERBAL PROBLEM SOLVING SITUATION

by Walter A. Watman

This investigation was designed to compare the effectiveness of programmed and traditional methods of instruction in mediating solutions to complex verbal problems. It was also concerned with the effect of changing an order of hint presentation.

A verbal problem that would minimize the contaminating effects of familiarity and previous experience yet not permit a quick and easy solution was selected, namely, the Denny Doodlebug Problem. Deceivingly simple at first glance, the problem requires for its solution the dismissal of certain beliefs commonly held by the subject and the adoption and synthesis of new ones.

The spirit of the research was basically exploratory and in keeping with this, there were no hypotheses regarding the independent variables manipulated. Four groups of subjects were utilized, each totaling approximately 20 subjects, both male and female. All were introductory psychology students. Two groups attempted to learn the information necessary to solve a problem by the programmed method of instruction (programmed groups). One of these groups was given three hints in what may be called the "customary" order while the other received them in "reverse" order where the third hint (rationally considered to be the most difficult and the most important) came first. The other two groups attempted to learn the information by the traditional method of instruction (traditional groups). Here also one group received the hints in the "customary" order, the other in "reverse" order. A short test was devised and administered to all four groups to determine whether method of instruction would affect transfer.

Method of learning of the information did not differentiate the groups as measured by performance on the first problem, but the number of solutions to problem 2 by those Ss who learned by the traditional method of instruction was significantly greater than the number obtained by the Ss who learned by programmed instruction. One group which learned by programmed instruction showed significantly less transfer of learned principles than did traditional instruction groups. A "reversed" order of hint presentation helped decrease solution time significantly on problem 1, but was not significantly more effective than the "customary" order with respect to problem 2 in spite of a trend in this direction.

Approved________________________

Date___________________________

ACKNOWLEDGMENTS

The author wishes to thank the members of his committee, Drs. Denny, Ratner and Kremen, for their time and their helpful criticisms, and D. J. Zerbolio for his generous and penetrating statistical advice.

To the staff of the university library goes the writer's appreciation for their complete cooperation.

Finally, the author is indebted to his wife Adrienne, whose continuous if not always compassionate understanding contributed in no small way to the success of this effort.

TABLE OF CONTENTS

Page

INTRODUCTION 104

METHOD 106

RESULTS 113

DISCUSSION 117

CONCLUSIONS 120

BIBLIOGRAPHY 130

LIST OF TABLES

Table Page

1. The effects of hint order on solution or non-solution of problem 2, irrespective of method of instruction 117

2. The effects of method of instruction on solution or non-solution of problem 2, irrespective of hint order 118

3. Transfer paradigm--the relationship between performance on problems 1 and 2 as measured by solution or non-solution 121

4. Changes in the proportion of solvers from problem 1 to problem 2 in each of the four groups 115

5. Relationships in the data for problem 1 123

6. Relationships in the data for problem 2 123

7. Sample size 123

8. Four Mann-Whitney U Tests on the solvers and four on the solvers plus the non-solvers, comparing solution times and hint orders on problem 1 124

9. The relationship of method of instruction and hint order to solution or non-solution of problem 1 124

10. The relationship of method of instruction and hint order to solution or non-solution of problem 2 124

11. Traditional subjects--hints in "customary" order (TO_1) 125

12. Traditional subjects--hints in "reversed" order (TO_2) 125

13. Programmed subjects--hints in "customary" order (PO_1) 126

14. Programmed subjects--hints in "reversed" order (PO_2) 126

LIST OF APPENDICES

Appendix Page

A. SOME CHANGES IN THE METHODOLOGY USED IN PREVIOUS RESEARCH AND WHY THEY WERE MADE 121

B. SOME TABLES OF RESULTS 123

C. TABLES SUMMARIZING RAW DATA 125

D. THE TWELVE QUESTION MULTIPLE CHOICE TEST 127

E. THE LINEAR PROGRAM 129

INTRODUCTION

This research attempted to examine the effectiveness of two different methods of instruction as they related to the solution or non-solution of a complex verbal problem solving task. Specifically, the experimenter tried to determine how well some conceptual material would be learned using one type of instruction as opposed to the other and which method fostered the best "use" of the concepts learned as measured by the solutions obtained to a complex verbal problem.

What is referred to in this study as the "traditional" method of instruction consists of the presentation to the subject of any sequence of information identical in appearance to a page taken directly from any standard text. In other words, it is the way by which most individuals obtain formal information. Programmed instruction, the second method, involves arranging some subject matter in sequential form according to certain established principles evolved from the psychology of learning. It acknowledges the importance of individual differences by beginning where the learner is and by permitting him to proceed at his own pace, and it requires that the learner be active, thus demanding his participation. Programmed learning emphasizes the organized nature of knowledge because it requires continuity between the easier (earlier) concepts and the harder (later) ones. It also provides immediate knowledge of results, the value of which, it may be added, is still debated (Hilgard, 1961).

A "program" consists of subject matter composed of sequential items which involves controlled presentation of material, active response of the learner (Skinner, 1958), the use of cues (prompts) to elicit correct responses and immediate confirmation of success or failure (feedback) which permits individual learners to progress independently and at their own pace (Homme and Glaser, 1959, page 104). In short, a program consists of material that is carefully worked out, well composed in terms of permitting the learner to be successful a high percentage of the time and can be placed in a "teaching machine" which may be any device that serves as a vehicle for the presentation of instructional material. Although commercial enterprises have produced many sophisticated teaching machines, all have certain basic characteristics in common (AV Communication Review, Volume 8, Number 2): (1) They all present a unit of verbal or symbolic information; (2) They provide the learner with some means of responding to each unit, by writing in most cases; (3) They inform the learner as to the correctness of his response. The major function of all these devices is to cause the learner to omit each response and to lead him through a great many steps devised to minimize the probability of an incorrect response. A linear program devised by M. R. Denny, which was a variation of the "programmed text" or paper teaching machine, was used in this research and is described in the Procedure section of this paper.

To investigate the effectiveness of programmed and traditional instruction, a complex verbal problem which was not amenable to simple rote memorization had to be found. The Denny Doodlebug problem (Rokeach, 1960, page 171), an extremely difficult verbal problem, was chosen for this purpose. It requires for its solution not only the overcoming of three currently held beliefs but also the adoption and synthesis of new ones. Research in the problem solving area necessitates determing if differences in the intelligence of the Ss could account for results obtained. In previous experimentation using the Doodlebug, Rokeach (1960, page 140) found that "correlations are generally small (-.36) between intelligence and ability to solve the Doodlebug problem."

Gagne and Smith (1962) revealed that Ss who verbalized, i.e., stated aloud why they were making a particular move, were significantly more successful than Ss who did not in terms of the number of moves necessary to solve the problem. Previous research, again by Rokeach (1960, page 171) using the Doodlebug, has also indicated that verbalizing questions and statements concerning the nature of the Doodlebug facilitated solutions. This will be discussed at greater length but for the present suffice it to say that this research supports these conclusions. Recent work by Gagne and Brown (1961) vital to the findings of the present experiment indicates that one learning program emphasizing a Discovery method analogous to Katona's learning of meaningful organization or learning with understanding (Katona, 1940) was more effective than a Rule and Example method analogous to a Skinner type small step program. Effectiveness was measured in terms of the transfer of what was learned to a problem solving situation. The transfer measure used indicated that "the programs containing small steps differed most in their effectiveness whereas the large step program fell in between." In view of this Gagne suggests that "what is learned," i.e., the mediators of the behavior, rather than "how it is learned," i.e., the behavior itself, is the crucial factor in learning effectiveness.

Evans _et al_. (in Lumsdaine, 1960, page 450) conducted an experiment with college students to compare the effectiveness of a sequence of information taken directly from a standard textbook presented in programmed form to one group and in traditional textbook form to another. The results indicated that subjects who interacted with the programmed text had a higher mean performance score on an independently constructed multiple choice test, but the difference was not significant. These subjects exhibited significantly less variability than did the Ss who received the traditional presentation of the same material.

The results of a pilot study suggested some changes in the methodology used in previous research with the Doodlebug. Explanations of the changes and why they were made are discussed in Appendix A.

METHOD

Subjects

A notice was posted near the introductory psychology classrooms indicating that subjects were needed for a psychological experiment and that research credit would be given. Both males and females were eligible. Such a notice was posted on two occasions, thus creating two different groups of approximately 20 Ss each. Two other groups of about the same number were obtained by requesting an introductory psychology instructor to ask her class for volunteers. All Ss were told that the experiment would take approximately two hours in spite of the fact that only one hour and fifteen minutes was needed. This was done to permit the Ss to devote their time to the tasks at hand instead of thinking about leaving the testing situation when the end of a stated time period appeared close at hand. All Ss were offered research credit and told that they need not participate if they did not wish to. All four groups contained about the same number of males as females.

To facilitate discussion let programmed instruction be designated by P and the groups receiving it be known as programmed groups. Traditional instruction shall be designated as T and the groups receiving it called traditional groups. One order of hint presentation shall be called the customary order and designated as O_1 while the other will be known as the reverse order and designated as O_2. The first group tested attempted to learn the information necessary to solve the Doodlebug problems by the traditional method of instruction (T) and were given the hints in reverse order (O_2). The second group learned the information by the programmed method of instruction (P) with the hints also given in reverse order (O_2). The third and fourth groups learned the information by the programmed and traditional methods respectively, and received the hints in customary order (O_1). Each group was tested on separate days. At one point during the experiment and at its end, the Ss were cautioned not to divulge or discuss either the solutions to the problems or any information they had acquired during the experiment.

Materials

The information to be learned concerned the way in which Joe Doodlebug functioned and was administered to the Ss by either the traditional or programmed method of instruction.* In appearance, the former consisted of one sheet of paper titled "Joe Doodlebug" with all the information mimeographed on it in the form illustrated below. All Ss were given the statement of the problem (described later in this chapter) along with the information to be learned.

---

*The average time taken to learn the information by the traditional method of instruction was 3.3 minutes and by the programmed method, 10 minutes (Appendix A). All Ss had 15 minutes to learn the information by either method in the actual experimental procedure. At the end of this time period, the information learned was removed but the Ss were permitted to retain the statement of the problem. It is possible then that the traditional groups were able to rehearse more often than the programmed groups during the 15 minute learning period; however, the programmed Ss were observed going over the program several times.

JOE DOODLEBUG

> Joe Doodlebug is a strange sort of imaginary bug who lives in a world where he can do certain things and cannot do others. He can jump in only four different directions: North, South, East, or West. He cannot jump diagonally: e.g., Southwest, Northeast, etc.
>
> Once Joe starts in any direction, that is, North, South, East, or West, he must jump four times in that same direction before he can switch to another direction.
>
> Joe can only jump; he cannot crawl, fly or walk. Joe can jump large distances or very small distances. Joe cannot turn around.
>
> Joe has been jumping all over the place getting some exercise when his master places a pile of food three feet directly West of him. Joe notices that the pile of food is a little larger than he. As soon as he sees the food, he stops dead in his tracks facing North.
>
> After all his exercise Joe is very hungry and wants to get to the food as quickly as he possibly can. Joe examines the situation and feels frustrated when he realizes that he will absolutely *have* to make four jumps and no fewer before he can get the food.

This same information presented by the programmed method consisted of a booklet of 53 sheets of paper arranged so that one page was labeled TEXT, the next QUESTION PAGE, and the following one, ANSWER PAGE (See Appendix E for the programmed material used). A sample of the first three pages of the booklet follows.

Page 1 TEXT

> Joe Doodlebug is a strange sort of imaginary bug who lives in a world where he can do certain things and cannot do others.

Page 2 QUESTION PAGE

> Joe is a st_____ bug who cannot do everything he wants to do.

Page 3 ANSWER PAGE

> Joe Doodlebug is a *strange* bug who cannot do everything he wants to do.

The first unnumbered page of the booklet contained the directions and read as follows:

> Please learn the information in this booklet. It is necessary to learn it well and understand it thoroughly in order to solve the problem. To help you do this, some pages are labeled TEXT and tell you what you need to know to solve the problem. Each text page is followed by one or more question pages (QUESTION PAGE). Please write the answers to these questions before turning to the next page where the correct answers can be found (ANSWER PAGE). If after seeing the correct answer you would like to check the text, it is all right to do so. Take each page as it comes and do not look ahead.

Problem 1 was inscribed on a single sheet of paper and given to all subjects along with the information to be learned. It read as follows:

> Since you now know all the things that Joe can and cannot do and you also know that if Joe could get to the food in fewer than 4 jumps he would, the problem is to explain why Joe Doodlebug takes exactly four jumps, no more and no less, in order to get the food.

The solution to the problem is as follows: At the moment Joe's master placed the food down, Joe had already jumped once to the east. He therefore has to jump sideways three more times to the east, and once sideways back to the west, landing on top of the food. He can now eat.

Problem 2, also stated on a single sheet of paper, read as follows:

> On another day when Joe is jumping about, his master places the food three feet directly north of him. Everything you have learned about Joe's abilities and limitations is exactly the same. Examining the situation, Joe finds that this time he will absolutely have to make three jumps and no fewer before he can get the food.
>
> The problem is to explain why this time Joe Doodlebug takes exactly three jumps, no more, no less, in order to get the food.

The solution to problem 2 is that at the moment Joe's master placed the food down, Joe had already jumped twice to the south. He therefore has to jump twice more to the south to complete his sequence of four jumps and once forward back to the north, landing on top of the food.

To infer the effectiveness of the methods of instruction all four groups attempted to solve this second problem which was identical for all groups and which required for its solution the transfer of learned principles dealing with how Joe Doodlebug functions.

While the Ss were attempting to solve problem 1, hints were displayed in the front of the room under conditions to be described. These hints were presented at the same time intervals to all groups and were not present when the Ss were to solve problem 2. Each of the three hints was printed in 3-inch red block letters on a piece of oaktag 18 x 48 inches and taped onto the blackboard at predefined time intervals. Hint 3 read as follows:

> When Joe sees the food and stops dead in his tracks facing north, he has already taken one jump in one of the four directions.

Hint 2:

> Joe can jump sideways and backwards as well as forward.

Hint 1:

> Joe does not have to face the food in order to eat it.

The presentation of the hints in this order, i.e., 3, 2, 1 respectively, was the reverse order of hint presentation (C_2) spoken of earlier. One programmed group and one traditional group received the hints in this order while the other two groups received the hints in the customary order (C_1), hints 1, 2, and 3 respectively.

In an attempt to obtain some idea of what bits of information could be related to a subject's solution or non-solution of problem 1, a short test of 12 multiple choice questions incorporating the points that the experimenter felt had to be learned in order to solve the problem was administered to all Ss immediately after the time period allotted to the learning of the information ended. Both forms A and B consisted of the same questions in different order and one copy of form B may be found in Appendix D.

Neither form was pretested for reliability or validity, but both were checked for ambiguity of wording and content. The 12 questions proved to be nondiscriminatory in that whether or not the subjects solved problem 1, nearly all obtained perfect scores on this test.*

---

*Daugherty found that the traditional groups learned and retained significantly more of the information necessary to solve problem 1 than did programmed groups as measured by this test.

Procedure

The research was carried on in a large introductory psychology classroom where the subjects were seated around the perimeter of the room to minimize the disturbance of neighbors when one S arose to offer the experimenter his solution. Four groups were used. Two attempted to learn the information by the programmed method of instruction and two by the traditional method. One programmed group received the hints in customary order, the other in reverse order. The same procedure was followed for the two traditional groups. The four groups may be referred to, then, as the TO_2 group, the PO_2 group, the TO_1 group and the PO_1 group. The four groups were tested on separate days in the above order and contained 17, 17, 16 and 17 members respectively.

Accuracy in the timing of a solution or an attempt at one was accomplished in the following manner. Another experimenter as familiar with the proceedings as this researcher, stood at the front of the room with him. Both possessed stop watches. Each subject was given a number before the experiment began and the group was told that "when you are sure you know the correct solution to a problem you will be given, walk quietly to the front of the room, mention your number, and tell the solution to one of us." Whenever an S arose the time was noted next to the Ss number. If the solution offered was incorrect, the time was crossed off. Running time was also kept to enable the experimenter to start and stop the various phases of the experiment.

In order for the solution offered to be correct the S had to state that Joe had already jumped once to the east and therefore must jump three more times in this direction and once back to the west, landing on top of the food. Similarly an acceptable solution for problem 2 required the S to state that Joe had already jumped twice to the south and therefore must jump two more times in this direction and once back to the north.

After the numbers were assigned the following instructions were read:

> Let me have your attention. Please do not talk to one another during the experiment. When you speak to one of the experimenters do so very quietly. Do not leave the room until you are told that you may do so. This is not a demonstration. This is an actual research project on problem solving. The problem you are to be given is difficult, but it is definitely solvable. We are interested in your finding the solution. The material you will be given contains all the information necessary to obtain the solution. Therefore it should not be necessary to ask questions. (Not one question was asked by any S.*) You will be given sufficient time to learn the

---

*Appendix A

> information thoroughly and you may review it as often as you like. After a period of time the information will be taken away from you and you will be tested on it. As already stated, the problem is a difficult one and it cannot be over-emphasized that your ability to solve it will depend entirely on how well you learn and remember all the information given to you. In a moment you will be given the information to be learned and the problem. Do not turn them over until instructed to do so.

Depending upon the group, the information in programmed form and problem 1 or the information in traditional form and problem 1 were placed in front of the subject face down. The experimenter then said:

> You should use this period of time to learn the information. In the next time period you should try and solve the problem. However, if at any time after reading and thinking about the problem you are sure that you know the correct solution, walk quietly to the front of the room and tell it to one of us. If you are told that your solution is incorrect, this does not necessarily mean that all your reasoning is wrong.** Return to your seat and continue trying to find the correct solution. Begin.

All subjects in each of the four groups were then given fifteen minutes during which they were to learn the information given them by either of the two methods of instruction. At the end of this period of time the experimenter said: "The time is up. Turn the material face down." The information was then taken from each subject, but the problem was left with them, face down. The short tests were then distributed face down and the Ss were told: "Put your name and number on these tests. Circle the correct answer." They were then given three minutes for the tests which were collected individually. Once again the experimenter repeated:

> If at any time after reading and thinking about the problem you are sure that you know the correct solution, walk quietly to the front of the room and tell it to one of us. If you are told that your solution is incorrect, this does not necessarily mean that all of your reasoning is wrong. Return to your seat and continue trying to find the correct solution. Here is a hint.

With this the other experimenter taped either hint 1 or hint 3 on the blackboard, depending on whether O_1 or O_2 was being used. Five minutes later the experimenter said: "Here is another hint." This was taped alongside the first. Five minutes later the experimenter announced: "Here is the third hint." This was taped next to the other two. Fifteen minutes later the experimenter said: "The time is up. Turn the problem face down. You may take a three minute

---

** Appendix A

break. Under no circumstances should you discuss the problem or anythings related to it with one another."

Upon the subject's return, all three hints had been removed from the blackboard and problem 1 had been collected. Problem 2 was then given to each subject face down. The researcher then said: "This will not take as long. Begin." The subjects were given four minutes to solve this problem. This time, all that was recorded was whether or not the S solved, i.e., a check was placed next to his number or a blank space was left. No solution time was recorded. At the end of the four minute period the experimenter said: "The time is up. If any of you now realize that you have seen or knew the solution to this problem before, please come up and tell me. You will be given credit regardless. Thank you for your cooperation." The entire group was then reminded not to discuss the problems with anyone and then dismissed.

Schematically then, the procedure was as follows:

1. The statement of the problem and depending on the group, the traditional or programmed form of the information to be learned, was given to each S face down.

2. Appropriate instructions were read and the Ss were allotted 15 minutes to learn the information.

3. When the time was up the information was collected and the problem was left with the Ss face down. The 12 question multiple choice tests were distributed. The Ss were given 3 minutes to complete them and then they were collected.

4. A portion of the instructions was repeated and then, depending upon whether the "customary" (O_1) or the "reversed" (O_2) order of hint presentation was being used, hint 1 or 3 was taped onto the blackboard in front of the room. Five minutes later the second hint was taped up followed in five minutes by the last.

5. Fifteen minutes later the Ss were told that the time was up, were reminded not to discuss the problem with anyone, and were given a 3 minute break.

6. Upon the Ss return all three hints had been removed and problem 1 had been collected. Problem 2 was given to each S face down and all Ss were given 4 minutes to solve it. When the time was up the Ss were asked to tell the E if they already knew the solutions or were familiar with any part of the experiment. They were assured that credit would still be given, reminded not to discuss the experiment and then dismissed.

The groups were tested on separate days in the following order: (1) TO_2 (2) PO_2 (3) PO_1 (4) TO_1.

RESULTS

Tables 5 and 6 in Appendix B indicate as succinctly as possible the various combinations of method of instruction and hint orders for the solvers and non-solvers of problems 1 and 2 respectively. Table 7 shows the size of the sample used for each of the four groups and the total number of subjects in the entire study. All statistical tests in this research were two-tailed and the confidence level used was .05.

The time taken to solve problem 1 was recorded for all Ss and, due to the truncated nature of the distribution, i.e., many Ss did not solve the problem in the allotted time, the differences between learning times under respective hint orders was tested [illegible] eight Mann-Whitney U tests (Siegel, 1956, page 110). Four utilized only those Ss who solved the problem and four others encompassed both solvers and non-solvers. It can be seen from Table B, also in Appendix B, that none of the groups differ significantly in terms of the time taken to solve, but of the solvers, the two groups that learned by programmed instruction and received the hints in different order (comparison 1) do differ significantly. ($U = 9$; $p < .028$; $n_1 = 7$; $n_2 = 8$) This suggests that the reversed order of hint presentation (O_2) yields faster solutions than does the customary order (O_1) for the programmed groups.

The interpretation of the relationship between method of instruction under both hint orders and solution or non-solution of problems 1 and 2 was made by a multinomial test. Table 9, Appendix B, suggests that for problem 1 no combination of type of instruction and hint order was more effective than any other combination in terms of the number of solutions yielded ($X^2 = 4.794$; $df = 3$; $p < .20$). Table 10, on the other hand, indicates that some combinations of instructions and hint order differ significantly from others and are more effective in producing solutions ($X^2 = 12.8$; $df = 3$; $p < .005$).

The non-significant results obtained, as indicated in Table 9, obviate the need for determining the effects of either hint order or method of instruction on the solution or non-solution of problem 1. With respect to problem 2, however, these tests are necessary. It can be seen from Table 1, page 11, that the data do not support the hypothesis that solution or non-solution of problem 2 and order of hint presentation are related ($X^2 = 2.021$; $p < .20$). The data do suggest that the reversed order (O_2) might possibly have been more effective if the total sample was larger and the cell proportions remained constant. Table 2, page 11, indicates clearly that solution or non-solution of problem 2 is dependent upon the type of instruction received. The number of solutions obtained by those SS who learned by the traditional method of instruction was significantly greater than the number obtained by the Ss who used programmed instruction ($X^2 = 8.628$; $p < .005$).

Table 1.--The effects of hint order on solution or non- solution of problem 2 irrespective of method of instruction.

	O_1	O_2	
Solution	9	16	25
Non-Solution	24	18	42
Totals	33	34	67

Table 2.--The effects of method of instruction on solution or non-solution of problem 2, irrespective of hint order.

	P	T	
Solution	6	19	25
Non-Solution	27	15	42
Totals	33	34	67

Table 3.--Transfer paradigm--the relationship between performance on problems 1 and 2 as measured by solution or non-solution.

		Problem 2		
		Solution	Non-Solution	
Problem	Solution	25	11	36
	Non-Solution	2	29	31
		27	40	67

Table 4.--Changes in the proportion of solvers from problem 1 to problem 2 in each of the four groups.

Group	Proportion of Solvers of Problem 1	Proportion of Solvers of Problem 2	N	Z
PO_1	.438	.0625	16	2.05 Sig.
PO_2	.470	.294	17	.705 Non-Sig.
TO_1	.470	.470	17	.344 Non-Sig.
TO_2	.765	.648	17	.373 Non-Sig.

If there was a transfer of learned principles from problem 1 to problem 2, one would expect that Ss who solved the first problem would solve the second and that those who did not solve problem 1 would not solve problem 2. It can be seen from Table 3, page 13, that there were significant transfer effects, i.e., performance on problem 2 was dependent upon performance in problem 1 ($X^2 = 24.917$; $p < .001$). Although the transfer is not perfect, the data suggest that if one did not solve problem 1 he was not likely to solve problem 2, but, conversely, if one did solve problem 1 there was a good chance that he would solve problem 2.

Transfer was defined as no change in the proportion of solutions from problem 1 to problem 2. Significant differences between the proportions of solvers on both problems, then, indicated poor or no transfer of learned principles. A test of the differences between proportions using the number of solutions to both problems as a criterion (See Table 4, page 13) indicated that while the group receiving programmed instruction and the hints in customary order (PO_1 group) seemed to do as well as the other three groups on problem 1 (Table 9--no significant differences), it does significantly more poorly than the other groups on problem 2, the transfer task ($Z = 2.05$).

Although there appears to be this significant difference in the performance of this group (PO_1) on problems 1 and 2, the experimenter is cognizant of the fact that this distribution (PO_1) has been used twice before within comparisons using the entire sample (Tables 9 and 10). In effect, then, the significance level has been raised beyond the stated .05 (to three chances, not 1, in 20 of significance) and the probability of a Type 1 error was increased.

DISCUSSION

The data do not fully answer the question of the effectiveness of both the programmed and traditional methods of instruction in eventuating solutions to complex verbal problems. No evidence was found to support a claim of greater effectiveness for either of the two methods on problem 1 as measured by either time to solve (Table 8) or number of solutions obtained (Table 9).* The data do suggest that programmed instruction was in one group significantly less effective than traditional instruction as measured by the capability of both to bring about the transfer of learned principles from problem 1 to problem 2 (Tables 3 and 4).

One interpretation of the first mentioned results may be made in terms of the choice made by this experimenter not to permit verbalization on the part of the subjects. That is to say, the Ss were allowed no interaction with the researcher which might have permitted them to get feedback concerning the appropriateness of their current problem solving attempts.** Gagne and Smith (1962) found that Ss who failed to verbalize produced significantly less solutions to a particular problem than did Ss who stated aloud what they were doing. All provious research with the Doodlebug has permitted this verbal feedback and Rokeach (in a personal communication, January, 1962) deems this a necessary factor in orienting and motivating the Ss toward the goal of solution. His research indicates approximately 25 per cent non-solution of the Doodlebug problem while this experimentation yields 52 per cent non-solution. With respect, then, to the importance of verbalization for problem solving the results of this research are consistent with previous findings. While such a great percentage of non-solvers seems a heavy price to pay, greater control over the experimental situation was in this experimenter's opinion unquestionably attained.***

A second possible explanation for the failure of the method of instruction variable to differentiate the groups as measured by performance on problem 1 may have been the removal of the information to be learned from all Ss at the end of 15 minutes. The high percentage of non-solutions may also have been due to this factor.

The results yielded by the experimental variable of the order of hint presentation may be explained by carefully looking at the movement belief, one of the three beliefs that must be overcome in order to solve the Doodlebug problem. If the S overcomes this belief, he ordinarily reaches the conslusion that Joe may have stopped in the middle of a sequence of jumps rather than at the end of a sequence. This is the key to the solution, Rokeach (1960, page

---

*Daugherty reports similar findings with respect to problems 1 and 2.

**Appendix A.

***Appendix A.

173) found that "many Ss have difficulty (in solving) because they assume that Joe is at the end rather than possibly in the middle of a sequence." The customary order of hint presentation, O_1, (it was not previously known by this name), the one used in previous research, exposed the facing belief (hint 1), the direction and the movement beliefs (hints 2 and 3) to the Ss in that order. In this research the reversed order, O_2, exposed the movement belief first in an attempt to discover if this order was more effective and as indicated in Table 8 it was, for problem 1. The fact that O_2 was no more effective than O_1 in eventuating solutions to problem 2 cannot really be explained by this experimenter, but the data suggest (Table 1) a potentially greater effect of O_2 and possibly, if the total sample was larger and the cell proportions remained constant, significance may have resulted.

Table 10 reveals that some combination of method of instruction and order of hint presentation significantly influenced solution or non-solution of problem 2. Table 1 suggests that it was not hint order which brought this about while Table 2 indicates that the method of instruction was of significant import.

The demonstration of significant transfer effects (Table 3) and the finding that the PO_1 group (programmed) was significantly less effective in terms of ability to transfer what was learned in attempting to solve problem 1 to the solution of problem 2 necessitates an attempt to explain why the methods of instruction did not differ significantly in effectiveness as measured by performance on problem 1 but did do so as far as problem 2 was concerned. The non-occurrence of verbalization mentioned earlier seems to be a strong factor in explaining the results obtained in connection with problem 1, but another possible explanation may be that the programmed instruction induced as much learning of the material necessary to solve the problem as traditional instruction did, creating a kind of "balancing out" effect, thus accounting for the failure of either type of instruction to prove significantly more effective than the other (Tables 8 and 9).

With respect to the significantly less effective performance of the programmed group (PO_1) in terms of the transfer of learned principles from problem 1 to problem 2, a plausible explanation seems to be that the programmed instruction may have facilitated learning which was highly specific to the task at hand but which "may not be capable of conveying 'understanding' in the sense of inducing transfer to new problem situations." (Gagne and Brown, 1961)

An examination of representative published programs made by Gagne and Brown (1961) indicated that "they (the programs) appear to be concerned primarily with the usage of words in a variety of stimulus contexts and it seems possible, therefore, for such programming of conceptual material to induce the learning of "verbally stated principles" which Katona (1940) considers to be less than optimal for either problem solution or transfer to problem solving situations." Whether or not this is the case in the present research cannot be told from present evidence. But a study designed to investigate the effects of systematic changes in the programming of conceptual learning materials as they bore upon later performance in a problem solving situation (Gagne and Brown, 1961) indicated best performance for a "Discovery" condition and worst for a "Rule and Example" condition. The former, it was suggested, was more effective because it required the Ss to actively produce (i.e., discover themselves) concepts to be used subseouently in the solving of new problems. This feature may have been lacking in the Rule and Example program.

This researcher is suggesting then that the original learning of the information concerning Joe Doodlebug by the traditional method of instruction may have required the Ss to actively produce concepts, i.e., "discover" them, and in so doing, facilitated the transfer of these learned principles to the second problem solving situation. The fact that there was no evidence in favor of greater effectiveness for either method of instruction on problem 1 and that programmed instruction was found to be significantly less effective than traditionsl instruction in terms of transfer seems to lend plausibility to the hypothesis that programmed instruction did foster learning which was highly specific to the problem at hand.

Left unanswered is the question of why the other programmed group (PO_2) was not also significantly less effective than the traditional groups in terms of transfer. As mentioned earlier, Rokeach (1960, page 173) has found that it is the overcoming of the movement belief, the one concerned with whether Joe was in the middle or at the end of a sequence of jumps, that was crucial to the solution of the problem. The reversed order of hint presentation (O_2) exposed the hint necessary to overcome this belief first, and perhaps in so doing provided the Ss in the PO_2 group with enough of the reasoning needed to perform well on problem 2. A task of future research would certainly be to describe with authoritative specificity exactly what the variations in these stimulus materials were.

CONCLUSIONS

It is possible to state some conclusions based on results obtained in this research as follows:

1. Neither method of instruction, programmed or traditional, was shown to be more effective than the other as measured by either solution time or number of solutions obtained to problem 1. The reversed order of hint presentation (O_2) yielded faster solution of problem 1 than did the customary order (O_1), for the programmed group.

2. The traditional method of instruction was inferred to be more effective than the programmed method as measured by the significantly greater number of solutions to problem 2 obtained by the Ss who learned by the former procedure. No evidence was found to indicate greater effectiveness of either hint order as measured by the solution or non-solution of problem 2, although the data suggested a trend favoring O_2.

3. In terms of the ability of either method of instruction to foster the transfer of learned principles from problem 1 to problem 2, one programmed group (PO_1) did significantly more poorly than both groups which learned by the traditional method of instruction.

APPENDIX A

SOME CHANGES IN THE METHODOLOGY USED IN PREVIOUS RESEARCH AND WHY THEY WERE MADE

Whenever the Doodlebug problem has been used in previous research, it has always been administered on an individual subject basis. The experimenter in all cases gave the S "feedback," i.e., when the S volunteered a solution or asked a question and these were wrong or misleading, the experimenter would correct him and point out what was wrong with the statement made. The running of some pilot Ss revealed that this was an important procedure, but that there were marked differences in the benefits that accrued to each S as a result of such feedback. To control the possibility of unequal benefits, the Ss in this research were told only that they were right or wrong and nothing else when they attempted a solution. Also disclosed were severe differences in the time taken to learn the information necessary to solve the problem using either the programmed or traditional methods of instruction. The term "learn" is used advisedly here, for this is what the Ss should have been doing during the allotted time period. There were severe differences then in the amount of time it took each S to tell the experimenter that he felt he had assimilated the necessary information. The mean time taken by Ss given the information in programmed form was 10 minutes and for those given it in traditional form 3.3 minutes.

It thus became apparent that if the effects of the independent variable (method of instruction) was to be reliably demonstrated, at least two variables had to be controlled. First, the possibility that the experimenter by giving feedback was contributing to the learning that the S alone should have been doing had to be eliminated. In instituting this control the researcher paid the penalty of decreasing the total number of solutions. Second, the time taken by both programmed and traditional subjects to learn the information had to be equated, something not done in previous research.

For the first time Joe Doodlebug was administered to a group. This resulted in obviating some methodological difficulties but created others. A drastic reduction of experimental time was accomplished. The experimenter as a variable was eliminated to a great extent for now unlike individual administration, all Ss in each group were presented with the same experimenter at the same time. Six of the ten pilot subjects failed to solve the problem when the experimenter said "no" and nothing else to a solution attempt. Questioning of the Ss revealed that the experimenter was misleading the Ss by such a response for not everything the S said was incorrect. To prevent this and still not divulge information that might contaminate the effects of the independent variable, the instructions "if you are told that your solution is incorrect, this does not necessarily mean that all your reasoning is wrong" were added to the actual experimental procedure. The possibility that in a group situation a question might be asked which would not only distract the group but might provide information about the problem was eliminated by telling the Ss that they would "possess all the information necessary to obtain the solution and so it should not be necessary to ask questions." None were asked.

The customary order of hint presentation has been the one used in all previous research with the Doodlebug. This alone prompted the experimenter to ask what would happen if the order were reversed. The inclusion of this experimental variable was also suggested when it was noticed that four solutions obtained by the pilot Ss came shortly after reading hint 3 and were usually accompanied by an illuminative expression like, "Of course, now I see." Prior research did not indicate whether more solutions were obtained after the presentation of this hint than after any other.

APPENDIX B

SOME TABLES OF RESULTS

Table 5.--Relationships in the data for problem 1.

	Solvers				Non-Solvers		
	O_1	O_2	Totals		O_1	O_2	Totals
P	7	8	15	P	9	9	18
T	8	13	21	T	9	4	13
Totals	15	21	N = 36	Totals	18	13	N = 31

Table 6.--Relationships in the data for problem 2.

	Solvers				Non-Solvers		
	O_1	O_2	Totals		O_1	O_2	Totals
P	1	5	6	P	15	12	27
T	8	11	19	T	9	6	15
Totals	9	16	N = 25	Totals	24	18	N = 42

Table 7.--Sample size.

Group	Number
PO_1	16
PO_2	17
TO_1	17
TO_2	17

Table 8.--Four Mann-Whitney U Tests on the solvers and four on the solvers plus the non-solvers, comparing solution times and hint orders on problem 1.

	Solvers				Solvers + Non-Solvers			
	U Obtained	U Critical	P		U Obtained	U Critical	P	
PO_1 vs. PO_2	9	--	.028	Sig.	112.5	81	--	Non-Sig.
TO_1 vs. TO_2	42	24	--	Non-Sig.	92	87	--	Non-Sig.
PO_2 vs. TO_2	45	24	--	Non-Sig.	109	87	--	Non-Sig.
PO_1 vs. TO_1	17	--	.232	Non-Sig.	120.5	81	--	Non-Sig.

Table 9.--The relationship of method of instruction and hint order to solution or non-solution of problem 1.

	TO_1	TO_2	PO_1	PO_2	
Solution	8	13	7	8	36
Non-Solution	9	4	9	9	31
Totals	17	17	16	17	67

Table 10.--The relationship of method of instruction and hint order to solution or non- solution of problem 2.

	TO_1	TO_2	PO_1	PO_2	
Solution	8	11	1	5	25
Non-Solution	9	6	15	12	42
Totals	17	17	16	17	67

APPENDIX C

TABLES SUMMARIZING RAW DATA

Table 11.--Traditional subjects--hints in "customary" order (TO_1)

Subjects	Time to Solve Problem 1 (Seconds)	No. of Hints Given	Solved Problem 2 (Yes or No)
1	541	0	Yes
2	1262	3	No
3	1555	3	Yes
4	1335	3	No
5	2340	3	Yes
6	1637	2	Yes
7	677	0	Yes
8	1980	3	Yes
9	No Solution	3	Yes
10	No Solution	3	Yes

$\bar{X} = 1415.8$

Range = 541 2340

Table 12.--Traditional subjects--hints in "reversed" order (TO_2)

Subjects	Time to Solve Problem 1 (Seconds)	No. of Hints Given	Solved Problem 2 (Yes or No)
1	1005	1	Yes
2	1580	3	Yes
3	1395	2	No
4	1550	3	Yes
5	1616	3	Yes
6	1150	1	Yes
7	1700	3	Yes
8	2075	3	Yes
9	973	1	Yes
10	1460	2	Yes
11	1340	2	No
12	1705	3	Yes
13	1515	3	Yes

$\bar{X} = 1466.4$

Range = 973 2075

Table 13.--Programmed subjects--hints in "customary" order (PO_1)

Subjects	Time to Solve Problem 1 (Seconds)	No. of Hints Given	Solved Problem 2 (Yes or No)
1	2151	3	Yes
2	2490	3	No
3	1720	3	No
4	1410	3	No
5	1880	3	No
6	1761	3	No
7	1560	3	No

$\bar{X}$ = 1853.1
Range = 1410 2490

Table 14.--Programmed subjects--hints in "reversed" order (PO_2)

Subjects	Time to Solve Problem 1 (Seconds)	No. of Hints Given	Solved Problem 2 (Yes or No)
1	1140	2	Yes
2	1500	3	No
3	1110	1	No
4	1145	1	No
5	1080	1	Yes
6	1945	3	Yes
7	815	0	Yes
8	1530	3	Yes

$\bar{X}$ = 1283.1
Range = 815 1945

APPENDIX D

THE TWELVE QUESTION MULTIPLE CHOICE TEST

Form B Name: . . . .

Directions: Circle the letter of the correct answer for each question.

1. After examining the situation, Joe realizes he will have to make ________ jumps and no fewer before he can get the food.

   (a) five
   (b) four
   (c) three
   (d) two

2. Because Joe is very hungry, he wants to get to the food

   (a) as slowly as he can
   (b) as quickly as he can
   (c) in two hours after he sees it
   (d) at the end of the day

3. Joe's master places the food directly ________ of him.

   (a) North
   (b) South
   (c) East
   (d) West

4. When Joe sees the food, he is facing

   (a) South
   (b) West
   (c) East
   (d) North

5. The pile of food is ____________

   (a) smaller than Joe
   (b) larger than Joe
   (c) the same size as Joe
   (d) green

6. Joe can jump

(a) not more than two inches at a time
(b) large or small distances
(c) not less than 10 feet per jump
(d) only 6 feet at a time

7. Joe can turn around

(a) never
(b) at the end of a sequence of four jumps in one direction
(c) in the air
(d) when he is hungry

8. Joe can only

(a) walk
(b) fly
(c) jump
(d) crawl

9. Joe can change directions

(a) at any time
(b) when he is moving south
(c) only after jumping four times in one direction
(d) when its a sunny day

10. Before Joe can switch to another direction, he must jump _________ times in the same direction.

(a) four
(b) three
(c) two
(d) one

11. Joe can jump diagonally

(a) at the end of a sequence
(b) never
(c) in the middle of a sequence
(d) when he feels like it

12. Joe can jump in only _______ differenct directions.

(a) one
(b) two
(c) three
(d) four

APPENDIX E

THE LINEAR PROGRAM

BIBLIOGRAPHY

Daugherty, M. 1962. Unpublished Master's thesis, Michigan State University.

Evans, J. L., Glaser, R. and Homme, L. E. 1960. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In Lumsdaine, A. A. and Glaser, R. Teaching Machines and Programmed Learning. Department of Audio-Visual Instruction, National Education Association.

Fillenbaum, S. and Jackman, J. 1961. Dogmatism and anxiety in relation to problem solving: An extension of Rokeach's results. J. Abnormal and Soc. Psych., 63, 212-214.

Fry, E. B., Bryan, G. L., and Rigney, S. W. Teaching machines: An annotated bibliography. AV Communication Review, 2, Supplement 1.

Gagne, R. M. and Brown, L. T. 1961. Some factors in the programming of conceptual learning. J. Exper. Psych., 63, 12-18.

Galanter, E. Automatic Teaching: The State of the Art. New York: John Wiley and Sons, Inc., 1959.

Hilgard, E. R. 1961. What Support from the psychology of learning. National Education Association Journal. November.

Homme, L. E. and Glaser, R. 1959. Relationships between the programmed textbook and teaching machines. In Galanter, E. Automatic Teaching: The State of the Art. New York: John Wiley and Sons, Inc.

Hoppe, R. H. 1960. Unpublished Master's thesis, Michigan State University.

Katona, G. 1940. Organizing and Memorizing. New York: Morningside Heights, Columbia University Press.

Lumsdaine, A. A. and Glaser, R. 1960. Teaching Machines and Programmed Learning. Department of Audio-Visual Instruction, National Education Association.

Rokeach, M. 1960. The Open and Closed Mind. New York: Basic Book Company.

Rokeach, M. 1962 Personal communication, January.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill Book Company.

Skinner, B. F. 1958. Teaching machines. *Science*, 128, 969-977.

Walker, H. and Lev, J. 1953. *Statistical Inference.* New York: Holt, Rinehart and Winston.

A COMPARISON OF THE EFFECTS OF PROGRAMMED AND CONVENTIONAL METHODS OF INSTRUCTION ON LEARNING AND PROBLEM SOLVING IN OLDER CHILDREN

By

Marilyn Daugherty

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1962

ABSTRACT

A COMPARISON OF THE EFFECTS OF PROGRAMMED AND CONVENTIONAL METHODS OF INSTRUCTION ON LEARNING AND PROBLEM SOLVING IN OLDER CHILDREN

Marilyn Daugherty

Michigan State University

This study was designed to compare the effects of a linear programmed method of instruction with a more conventional method when learning time, instructor, intelligence, information presented, and motivation factors were the same for both groups. "Conventional" material was in traditional printed narration form without accompanying lecture-type instruction. The same material in linear-type programmed form was given to the experimental group. I.Q. scores were obtained for each subject so groups could be equated on this variable. Learning criteria was a multiple-choice test. Problem-solving performance was reflected by correct solutions to problems presented and solving-time required.

Results showed significantly better learning for conventional subjects than program subjects. Programming of the material to be learned resulted in no better learning for subjects of below 100 I.Q. and resulted in poorer learning for subjects of above 100 I.Q. Type of instruction did not prove to be an influencing factor on whether or not subjects solved the problems presented or on solving-time required by those who did solve them.

The above results suggest that at least in some areas of learning and particularly for students of high-average or better intelligence, programmed instruction is still a very open question.

Approved ______________________
Major Professor

Date ______________________

ACKNOWLEDGMENTS

I wish to thank Dr. M. Ray Denny for his excellent guidance throughout this study. He also provided the linear program which was used.

The cooperation of the Grand Lodge, Michigan, and De Witt, Michigan, school system is also greatly appreciated. Without such help this study would have been impossible.

I also wish to thank Drs. Ratner and Dietze of the Michigan State University Psychology Faculty for their helpful suggestions concerning the writing of the text.

TABLE OF CONTENTS

Page

I. CONTEXT AND STATEMENT OF PROBLEM 132

II. METHOD

1. Subjects 135
2. Materials 135
3. Procedure 136

III. RESULTS

1. Learning 137
2. Problem-Solving 137

IV. DISCUSSION 138

V. SUMMARY AND CONCLUSIONS 143

FOOTNOTES 144

REFERENCES 146

APPENDICES 148

LIST OF TABLES

Page

I. Frequency of Multiple-Choice Test Scores by I.Q. 139

II. Frequency of Solutions to Problem I by I.Q. 140

III. Frequency of Solutions to Problem II by I.Q. 141

APPENDICES

Page

A. Linear Program 148

B. Conventional Presentation 149

C. Joe Doodlebug Problems I and II 150

D. Multiple-Choice Test 152

E. Verbal Instructions 153

F. Wilcoxon Test on Program and Conventional Ss' Multiple-Choice Test Scores 155

G. Significance Test on Program Ss' Scores and I.Q. Group 156

H. Significance Test on Conventional Ss' Test Scores and I.Q. Group 157

I. Significance Test of Above 100 I.Q. Ss and Perfect and Non-Perfect Test Scores . . . 158

J. Significance Test of Below 100 I.Q. Ss and Perfect and Non-Perfect Test Scores . . . 159

K. Significance Test of Program and Conventional Ss' Problem I Solutions 160

L. Significance Test of Program and Conventional Ss' Problem I and Problem II Solutions 161

M. Wilcoxon Test on Program and Conventional Ss' Problem I Solution Times 162

N. Significance Test on All Solvers and Non-Solvers of Problem I and I.Q. Group . . . 163

O. Significance Test on All Solvers and Non-Solvers of Problems I and II 164

A Comparison of the Effects of Programmed and Conventional Methods of Instruction on Learning and Problem Solving in Older Children

Marilyn Daugherty

Michigan State University

The basic idea in programming is that the most efficient, pleasant, and permanent learning occurs when the student proceeds through a course by a large number of small, easy-to-take steps.[1] A program is a sequence of carefully constructed items designed to lead the student to mastery of the subject with minimal error. Information is given to the student in small units to which he responds in some way. Items are designed so that the student can make correct responses while progressing toward more and more complex material.[2]

In the 1920's Sidney L. Pressey designed several machines for automatic testing. The student selects a multiple-choice answer. If wrong, the error is tallied and the student continues shoosing until correct; if right, the device moves to the next item. This "machine" was designed against a theoretical background of the principles of "frequency" and "recency"; "massed" and "spaced" practice.[3]

In recent years a "scrambled" type of programming has been designed whereby incorrect responses to multiple-choice items cause the device to switch to material designed to clarify misunderstanding and then proceed from there. In "linear" programming, the student first supplies the answer; he is then given the correct answer immediately and the device continues. B. F. Skinner has said that an important feature of efficient programming should be that the student must compose his response rather than select it from a set of alternatives.[4] One reason for this is that it is important for the student to recall rather than to recognize, and to make a response as well as see that it is right. Another reason is that effective multiple choice material must contain plausible wrong answers which are out of place in the delicate process of "shaping" behavior because they strengthen unwanted forms. According to Skinner, "The act of responding tends to cause learning; therefore the student should not be exposed to incorrect alternatives."[5] Linear programming is the type employed in the present study.

"Program subjects" referred to in this study are subjects who received the program found in Appendix A. The term "conventional" may apply to any of several traditional forms of teaching presentation including lecture, group discussion, and printed narration form. As used in the present study however, "conventional" refers to a printed narration form only. No lecture or discussion took place.

There are five advantages over other types of instruction claimed for programming: (1) If each step is small and easy to take, the student is not likely to make erros. If a student makes many errors in learning, he often decides that he does not like the subject. (2) Better learning occurs when the student is actively responding to the material. He "learns by doing." (3) Immediate feedback of correct responses results in better learning. (4) Each student can work each step as slowly or quickly as he chooses. Programs are self-paced. (5) By looking over a program, one can see exactly what came before a step on which a mistake was made. Since accurate records of the learning experience of each student are available, revisions can be made on the basis of actual student responses. If the presentation of some point is not clear, this will show up on the performance of the program.[6]

A study on automated teaching methods using linear programs by A. Roe (University of California) compared the effects of (a) teaching machines; programmed textbooks; programmed lectures; and standard lectures; (b) free response vs. multiple-choice responses; and (c) individual booth vs. classroom. He found that all program groups performed significantly better than the conventional lecture group. The free-response group took more time than the multiple-choice and programmed text groups but there were no significant differences between them when tested. There was no indication that any one of the methods was particularly better for a particular aptitude level on the basis of the test criterion or learning time.[7]

J. H. Hughes compared the effectiveness of programmed and conventional classroom instruction and found that program students performed significantly better than the conventional students.[8] Coulson and Silverman found similar results in a study on teaching machines.[9]

The relation of programming to problem-solving has scarcely been investigated. Watman (Michigan State University) found that significantly more college age subjects in the group receiving a traditional printed narration presentation of the materials solved the "Joe Doodlebug" problem correctly than program subjects.[10] Watman's study--particularly in regard to materials used and experimental controls employed--was very similar to the present study.

Douglas Porter conducted a 22 week experiment on the teaching of spelling to elementary school children using teaching machine instruction. The experimental (program) group received the same words as the control group who used standard textbooks. Results were measured by standardized achievement test scores. The experimental group was significantly superior. No relationship between intelligence scores and achievement in the experimental group appeared, but a significantly positive relationship was shown in the control group. Porter's experiment seems to indicate that this type of instruction is particularly good for the lower I.Q. groups.[11]

Much exploratory study and definitive experimentation is necessary to ascertain the merits and disadvantages of programming compared to more conventional types of instruction in a learning situation.[12] It seems that of accessible published studies involving comparisons of programmed and conventional instruction, it is almost impossible to find studies utilizing strict experimental controls. A common error seems to be that of introducing "teaching machines" or programmed materials to one group and then comparing performance of that group with another group who remains in its usual routine. Little effort seems to have been made to account for: (a) interaction between instructor and pupils in the conventional group; (b) interaction between experimental (program) subjects and experimenter (usually a new personality to this group); and (c) how similar or different the specific information used in program and conventional groups may be. For example, when "lecture" is a part of the conventional group's instruction, which of numerous possible methods of lecture was used and whether or not each point presented to the program group was also presented in the lecture group seems never to be mentioned. The relationship between programmed instruction and intelligence is another area inviting research.

The following study was designed to add one small drop of understanding to the--so far--very small pool of knowledge surrounding programmed instruction. An attempt was made to compare the effects of conventional and programmed methods of instruction on learning and problem-solving under the most stringent controls possible. The study was undertaken to find out whether or not subjects with conventional instruction would perform as well or better than program subjects when learning time, instructor, intelligence, information presented, and motivation factors (influenced by instructions, instructor, etc.) were the same for both groups. In order to help insure equal treatment of the two groups, the conventional subjects were given the material in traditional printed narration form. "Lecture" was not implied by the term "convnetional". The material thus appeared to the conventional subjects in the same way as much of their school subject-matter routinely appears in reading assignments. These subjects however, were taken out of their daily routine for the experiment as were the program subjects. Also, the material was given to them by the researcher--not by their regular classroom teacher. Thus both groups had a novel experience. This aspect of the study is rather important since it is quite possible that seeming superiority of programmed instruction may sometimes be due to a new and novel experience by the program group and not a result of the program itself.

Another question of interest in this study was whether or not programmed instruction would result in superior performance by students of low average or lower intelligence compared to the performance of subjects of similar intelligence receiving conventional instruction as a Porter's study previously mentioned.

Method

Subjects

One hundred fourteen (114) eighth grade students from 13 to 15 years of age were used (57 program and 57 conventional subjects). Approximately half of these were from the Jonas Sawdon Jr. High School of Grand Ledge, Michigan, and the other half from the De Witt Jr. High School, De Witt, Michigan. There were approximately half male and half female students in both the program and conventional groups. Both schools are located in small communities and enroll average students. Of the subjects in the program group participating in the study, the mean I.Q. was 100 and I.Q. scores ranged from 60 to 136. Of the subjects in the conventional group, the mean I.Q. was 100 and I.Q. scores ranged from 59 to 131.

Materials

I.Q. scores were obtained by administering the Otis Test of Mental Ability (intermediate);[13] and the Science Research Associates Primary Mental Abilities test (intermediate).[14]

Material to be learned was information about a hypothetical character named Joe Doodlebug. Both control and experimental groups received the same information but the experimental group received it in linear programmed form with the necessary additional printed instructions. The program was constructed by Dr. M. Ray Denny (Michigan State University). This program will be found in Appendix A. (See page 129).

The two problems faced by Joe Doodlebug and presented to the subjects to be solved will be found in Appendix C. (Problem I was presented to all subjects at the very beginning of the learning session along with the information to be learned).

A 12-item multiple-choice test over the information about Joe Doodlebug was used in two forms and distributed alternately to subjects. (Appendix D). This was used as a measure of how well the material had been learned.

Three hints which were given to the subjects were clearly printed on large posters and placed in front of the groups at the proper time intervals. (Appendix E).

Procedure

Students were assigned to the two groups on the basis of I.Q. scores. Of the two highest scorers, one was assigned to the conventional and the other to the program group; the next two highest scorers similarly, etc., until all were assigned. This was done prior to the day of presentation of the material to be learned and the problems. Each school was treated separately in this way.

Since all conditions for program and conventional subjects were the same except for method of printed presentation, it was possible to work with experimental and control subjects in the same learning session. Four 8th grade classes of approximately 30 students in each and each containing both conventional and program subjects were used (each S wore a tag with a number corresponding to his name on the researcher's record, and a symbol indicating whether he was in the experimental or control group). The exact verbal instructions will be found in Appendix E.

Verbal instructions were given. The information to be learned and Problem I was then distributed. Seventeen minutes were allowed to learn this material. During this learning period, both groups had Problem I available. Thus the purpose for learning the material and its importance in finding the correct solution to the problem was equally clear to both groups. The information was then taken away, the test administered and collected. The next 25 minute period was allowed for the subjects to offer solutions to the Joe Doodlebug Problem I. During this period, any subject wishing to offer a solution would move to the front of the room and quietly tell the researcher or assistant. If correct, he would return to his seat and wait quietly during the remainder of the period. A subject could offer as many solutions as he wished until he found the correct one. Each of three hints was given at 5 minute intervals at the beginning of this period. Problem I was then collected and a 3 or 4 minute break was given. Problem II was then distributed and 5 minutes were allowed for solutions to be offered. The researcher (and an assistant) wrote down the solution time when any subject offered the correct solution.

Results

Learning

(a) The 12-item multiple-choice test scores which were used as a measure to reflect learning as a function of type of instruction showed significantly better performance by conventional subjects than by program subjects. The Wilcoxon matched-pairs signed ranks test was used and a Z score of -2.33 derived which is significant at the .05 level. (Appendix F) There were 5 subjects in the experimental group who did not complete the program during the learning period. Four of these students had scored in the lower 1/5 of the group on the I.Q. tests. The other student achieved an I.Q. score in the average range. Table I on page 139 lists scores for both groups.

(b) No significant differences as to perfect or non-perfect test scores were found at the .05 level between above and below 100 I.Q. subjects in the program group. (Appendix G)

(c) Significant differences in test performance using perfect and non-perfect test scores were found between above and below 100 I.Q. subjects in the conventional group at the .05 significance level. (Appendix H) More high I.Q. subjects received perfect scores than low I.Q. subjects.

(d) Significant differences in perfect or non-perfect test scores were found between program and conventional subjects of above 100 I.Q. at the .01 significance level. (Appendix I) More conventional subjects received perfect scores than program subjects.

(e) No significant differences in test performance at the .05 significance level were found between program and conventional subjects of below 100 I.Q. (Appendix J)

The above results suggest that at least in some areas of learning the value of programmed instruction is still a very open question. Programming of the material to be learned resulted in no better learning for subjects of I.Q. below 100 and resulted in poorer learning for subjects of I.Q. above 100.

Problem-Solving

(a) No significant differences were found as to number of solutions between program and conventional groups on either Problem I (Appendix K) or Problem II (Appendix L) at the .05 level.

Tables II and III (pp. 140 and 141) show the number of solutions obtained for the two problems in each group.

(b) No significant differences as to solution time were found between program and conventional groups for subjects who did solve the problem. (Appendix M)

(c) Significant differences at the .01 level were found between above 100 I.Q. subjects and below 100 I.Q. subjects of both groups as to the number of problem solutions. The data indicates that the above 100 I.Q. subjects were much better able to solve than below 100 I.Q. subjects in both groups. (Appendix N)

(d) Highly significant differences between solvers and non-solvers of Problem I were found as to whether or not they solved Problem II. Almost all who solved I also solved II and very few who did not solve I solved II. (Appendix O)

Type of instruction did not prove to be an influencing factor on whether or not subjects solved the particular problems presented or on solving-time required by those who did solve them. Also there was no evidence of a relationship between how well the material was learned (as reflected by the multiple-choice test) and successful problem-solving. Studies by Rokeach indicate that certain personality factors--attitudes basic to the individual--account for most differences in solving this particular problem.[15]

Discussion

One consideration in attempting to account for the superior learning of conventional subjects is the fact that both conventional and program groups were given exactly the same amount of time in which to learn the information. Generally, studies reported on programmed instruction state that program groups take more time than other types of instruction but result in better learning. In this study however, program subjects had to use part of their learning period turning 52 pages and filling in 40 blanks while conventional subjects had all the information in narrative form on one page and were free to read and review it many more times than program subjects. If each conventional subject had been allowed to terminate the learning period at his own discretion, perhaps results would have been different. When superior learning seems to occur by program groups, it may be a function of spending more time and making more responses to the material and not a function of programming itself. More time and responding may have an adverse affect also according to David Cram who says: "Students tend to tire more quickly when working on linear programs . . . The pace is slower and more concentrated; there is more repetition; and the student is forced to make more

TABLE I

Multiple-Choice
Test Scores

	Conventional Ss									Program Ss								
I.Q.	12	11	10	9	8	7	6	5	4	12	11	10	9	8	7	6	5	4
130-139	1										1							
120-129	3	1								2	1							
110-119	8	1								5	4	1						
100-109	15	2								6	13	1	1					
90- 99	9	3	1							7	4		1					
80- 89	5		2								1			1	1		1	
70- 79	1	1					1					1	3			1		
60- 69								1	1	1								
59							1											

FREQUENCY OF TEST SCORES BY I.Q.

TABLE II

I.Q.	Conventional Ss Solved	Conventional Ss Non-solved	Program Ss Solved	Program Ss Non-solved
130-139	1	0	1	0
120-129	2	2	2	1
110-119	7	2	6	4
100-109	6	11	5	16
90- 99	1	12	4	8
80- 89	1	6	0	4
70- 79	1	2	0	5
60- 69	0	2	0	1
59	0	1		
TOTALS	19	38	18	39

FREQUENCY OF SOLUTIONS TO PROBLEM I BY I.Q.

TABLE III

I.Q.	Conventional Ss Solved	Conventional Ss Non-solved	Program Ss Solved	Program Ss Non-solved
130-139	1	0	1	0
120-129	3	1	2	1
110-119	6	3	5	5
100-109	4	13	6	15
90- 99	3	10	4	8
80- 89	1	6	1	3
70- 79	1	2	0	5
60- 69	0	2	0	1
59	0	1		
TOTALS	19	38	19	38

FREQUENCY OF SOLUTIONS TO PROBLEM II BY I.Q.

overt responses per unit of time."[16]

The comparatively poor learning evidenced by the experimental group could possibly be due to an inferior program. Possibly a different kind of program would be more beneficial than the linear program which was used, or this one improved upon. For example, if at the end of the program a page of summary were added so that the subjects could have all the information before them at once and allowing them a comprehensive review, better learning may result.

Another possible partial explanation for the results obtained on the multiple-choice test is that as each program subject progressed to the next step, he may have had a feeling of finality as to the previous step (having filled in the blanks successfully) and began to forget it. This is of course suggesting a phenomenon something like the Zeigarnik effect.[17]

Gagné and Smith conducted a study with ninth and tenth grade boys all above 110 I.Q. which may shed some light on the differences in performance of conventional and program subjects confronted with the "Joe Doodlebug" problem (although their study involved a problem different from the ones used in this study). They were interested in the effects of the subjects' own verbalizing during problem-solving performance. Results appeared to indicate that requiring subjects to verbalize during practice had the effect of making them think of new reasons for their moves, and thus facilitated both the discovery of general principles and their employment in solving successive problems.[18] Since conventional subjects needed less time to actually read and review the material, they were free during the remaining time to ponder the information and "verbalize" silently to themselves. It seems reasonable to suggest that (silent) verbalizing by the conventional subjects could partially explain their superior learning of the material in the light of Gagné and Smith's study.

Another question which seems plausible in regard to superior learning by program groups reported in some studies is whether the motivation factor increases for program subjects as a result of a new set of stimuli. If programmed instruction were continued over a long period of time, it is questionable whether or not the program group would continue to perform better than other types of instruction groups. The novelty of being in an experimental group and being presented with new materials to use was controlled in the present study by treating the control subjects in exactly the same manner as the program subjects. All subjects in both groups had their usual routine interrupted in order to participate in an experiment. All subjects worked with information quite different from usual school subject matter and with persons new to them. In this way the "novelty factor" which presumably operates in favor of superior performance by program groups in many studies was eliminated.

As for the "self-pacing" claim for programmed instruction, the conventional subjects in this study were actually more self-paced than the program subjects since the latter were required to fill in all the blanks and were thus "tied" to the program. Leslie Briggs conducted an experiment in which each of two groups was given a total of 13 minutes practice on a 20- item paired-associate problem. One group was self paced; pacing was automatic for the other. Both were then given a test trial. No differences between effectiveness of the two methods was demonstrated.[19]

It would seem from this study that students of high average or better intelligence may find programmed instruction (at least in some situations) too cumbersome to be advantageous. Certainly such investigation is a vital current need and requires more--and in general, better controlled--studies than are now available before the many questions surrounding programmed instruction can be answered.

Summary

This study was designed to compare the effects of a linear programmed method of instruction with a more conventional method when learning time, instructor, intelligence, information presented, and motivation factors were the same for both groups. "Conventional" material was in traditional printed narration form without accompanying lecture-type instruction. The same material in linear-type programmed form was given to the experimental group. I.Q. scores were obtained for each subject so groups could be equated on this variable. Learning criteria was a multiple-choice test. Problem-solving performance was reflected by correct solutions to problems presented and solving-time required.

Results showed significantly better learning for conventional subjects than program subjects of 8th grade level. Programming of the material to be learned resulted in no better learning for subjects of below 100 I.Q. and resulted in poorer learning for subjects of above 100 I.Q. Type of instruction did not prove to be an influencing factor on whether or not subjects solved the problems presented or on solving-time required by those who did solve them.

The above results suggest that at least in some areas of learning and particularly for students of high-average or better intelligence, programmed instruction is still a very open question.

Footnotes

[1] Slouth, Dewey A., "Descriptive Statistics" (Denison University), The Psychological Record, Vol. 10, pp. 224-226.

[2] Center for Programmed Instruction, Inc. (pamphlet--no author given), May, 1962.

[3] Markle, Eigen, and Komoski, A Programmed Primer on Programing, New York, 1961.

[4] Skinner, Cumulative Record, New York, 1959, p. 159.

[5] Cram, David, Explaining Teaching Machines and Programming, Fearon Publishers, San Francisco, 1961, p. 34.

[6] Principles of Programed Learning (No author given), 1961.

[7] Roe, A., et. al., "Automated Teaching Methods Using Linear Programs" (University of California), Programed Instruction, Vol. I, May 1961.

[8] Hughes, J. H., "The Effectiveness of Programmed Instruction: Experimental Findings for 7070 Training", Programed Instruction, Vol. I, New York, January 1961.

[9] Coulson and Silverman, "Effects of 3 Variables in a Teaching Machine", Journal of Educational Psychology, Vol. 51, pp. 135-143.

[10] Watman, Walt, "A Study of the Effectiveness of Programmed and Traditional Methods of Instruction in a Complex Verbal Problem- Solving Situation", Master's Thesis, Michigan State University, 1962.

[11] Porter, Douglas, "Some Effects of Year-Long Teaching Machine Instruction", Teaching Machines: An Annotated Bibliography, Department of Audio-Visual Instruction, Washington D.C., 1960, pp. 58-59.

[12] Center for Programed Instruction, op. cit.

[13] Otis Self-Administering Tests of Mental Ability, Arthur B. Otis, Intermediate Examination, Forms A and B, World Book Company, Chicago, 1922.

[14] SRA Primary Mental Abilities, L. L. and Thelma Thurstone (University of North Carolina), Science Research Associates, Inc., Chicago, 1948.

[15] Rokeach, Milton, The Open and Closed Mind, New York, 1960, pp. 171-196.

[16] Cram, op. cit., p. 52.

[17] Ruch, Floyd L., Psychology and Life, Scott, Foresman, and Co., Chicago, 1948, p. 379.

[18] Gagné and Smith, "Effects of Verbalization in Problem-Solving", Journal of Experimental Psychology, Vol. 63, p. 12.

[19] Briggs, Leslie J., et. al., "Self-Pacing Vs. Automatic Pacing of Practice on the Subject Matter Trainer", Teaching Machines: An Annotated Bibliography, op. cit., p. 24.

References

Center for Programmed Instruction, Inc. (pamphlet--no author given), May, 1962.

Coulson, John E., and Silverman, Harry F., "Effects of 3 Variables in a Teaching Machine", Journal of Educational Psychology, Vol. 51, June, 1960, pp. 135-143.

Cram, David, Explaining Teaching Machines and Programming, Fearon Publishers, San Francisco, 1961.

Fry, Edward B.; Bryan, Glenn L.; and Rigney, Joseph W.; Teaching Machines: An Annotated Bibliography, Department of Audio-Visual Instruction, Washington D. C., 1960.

Gagné and Smith, E. C., "Effects of Verbalization in Problem Solving", Journal of Experimental Psychology, Vol. 63, January, 1962, p. 12.

Hughes, J. H., "The Effectiveness of Programmed Instruction: Experimental Findings for 7070 Training", Programed Instruction, Vol. I, Bulletin No. 1, Issued by the Center for Programed Instruction, New York, January, 1961.

Markle, Susan M; Eigen, Lewis D.; and Komoski, P. Kenneth, A Programed Primer on Programing, Published by the Center for Programed Instruction, Inc., New York, 1961.

Principles of Programed Learning, (no author given), Teaching Machines Inc., New York, 1961.

Roe, A., et. al., "Automated Teaching Methods Using Linear Programs" (Department of Engineering, U. of California), Programed Instruction, Issued by the Center for Programed Instruction, Vol. I, Bulletin No. 1, May, 1961.

Rokeach, Milton, The Open and Closed Mind, Basic Books, Inc., New York, 1960.

Ruch, Floyd L., Psychology and Life, Scott, Foresman, and Co., Chicago, 1948.

Skinner, B. F., Cumulative Record, Appleton-Century-Crofts, Inc., New York, 1959.

Slough, Dewey A., "Descriptive Statistics" (Denison University), The Psychological Record, Vol. 10, pp. 224-226, Albuquerque, 1960.

Watman, Walt, "A Study of the Effectiveness of Programmed and Traditional Methods of Instruction in a Complex Verbal Problem-Solving Situation", Master's Thesis, Michigan State University, 1962.

APPENDIX A

LINEAR PROGRAM

(See Page 129)

APPENDIX B

CONVENTIONAL PRESENTATION

(See Page 107)

APPENDIX C

JOE DOODLEBUG PROBLEMS I AND II

THE PROBLEM:

Since you now know all the things that Joe can and cannot do and you also know that if Joe could get to the food in fewer than 4 jumps he would, the problem is to explain why Joe Doodlebug takes exactly four jumps, no more and no less, in order to get the food.

PROBLEM II:

On another day when Joe is jumping about, his master places the food three feet directly North of him. Everything you have learned about Joe's abilities and limitations is exactly the same. Examining the situation, Joe finds that this time he will absolutely have to make three jumps and no fewer before he can get the food.

The problem is to explain why this time Joe Doodlebug takes exactly three jumps, no more, no less, in order to get the food.

APPENDIX D

MULTIPLE-CHOICE TEST

(See Page 127)

APPENDIX E

VERBAL INSTRUCTIONS

"This is not a demonstration. This is an actual research project on problem solving. The puzzle you are to be given is difficult but it is definitely solvable. We are interested in your finding the solution. The material you will be given contains all the information necessary to obtain the solution. Therefore it should not be necessary to ask questions. You will be given sufficient time to learn the information thoroughly and you may review it as often as you like. After a period of time the information will be taken away and you will be tested over it. As already stated, the puzzle is a difficult one and it cannot be over-emphasized that your ability to solve it will depend entirely on how well you learn and remember all the information given to you. If you do not learn this information well, you will not be able to solve the puzzle. In a moment you will be given the information to be learned and the puzzle. You are not to turn them over until instructed to do so."

(Information and Problem I distributed).

"You should use this period of time to learn the information which you will need to solve the problem or puzzle. You will have a chance to tell one of us your answer in the next time period. If you think you have the answer, the best thing to do is to look back over the material and see if your answer is really right, and spend your time making sure you understand the information thoroughly. Begin." (17 minutes allowed). "The time is up." (Material collected. Tests passed). "Write your name and number. Circle the correct answer. Begin." (5 minutes allowed. Tests collected).

(Subjects retain copies of the problem). "In this time period you should try and solve the puzzle. Any time from now on that you are sure you know the correct solution, walk quietly to the front of the room and whisper it to one of us. Make sure you speak softly enough so no one else can hear you. If you are told your answer is wrong, this doesn't mean everything you thought is necessarily wrong; just return to your seat and continue trying to find the correct solution. You can have as many tries as you like. You may also write or scribble on the problem sheet you have there if it will help you. Here is a hint:"

> When Joe sees the food and stops dead in his tracks facing North, he has already taken one jump East. He must take 4 more jumps to get the food.

(5 minutes given). "Here is another hint:"

Joe can jump sideways and backwards as well as forward.

(5 minutes given). "Here is the last hint:"

Joe does not have to face the food in order to eat it.

(15 minutes given. Problems collected. 3 or 4 minute break. Problem II distributed). "This is the 2nd and last puzzle." (Same instructions as Problem I. 5 minutes given).

APPENDIX F

WILCOXON TEST ON PROGRAM AND CONVENTIONAL Ss' MULTIPLE-CHOICE TEST SCORES

$$Z = \frac{209.8 - \frac{38(39)}{4}}{\sqrt{\frac{(38)\ (39)\ (77)}{24}}} = -2.33$$

Significant at .05 level.

APPENDIX G

SIGNIFICANCE TEST ON PROGRAM Ss' SCORES AND I.Q. GROUP

Test Scores	Above 100 IQ	Below 100 IQ	
Perfect	13	8	21
Non-Perfect	22	14	36
	35	22	57

Z= -.21

Non-significant

APPENDIX H

SIGNIFICANCE TEST ON CONVENTIONAL Ss' SCORES AND I.Q. GROUP

Test Scores	Above 100 IQ	Below 100 IQ
Perfect	27	15
Non-Perfect	4	11

Z= 2.2

Significant at .05 level.

APPENDIX I

SIGNIFICANCE TEST OF ABOVE 100 I.Q. Ss AND PERFECT AND NON-PERFECT TEST SCORES

Test Scores	Program Ss	Conventional Ss	
Perfect	13	27	40
Non-Perfect	22	4	26
	35	31	66

Z= 3.89

Significant at .01 level.

APPENDIX J

SIGNIFICANCE TEST OF BELOW 100 I.Q. Ss AND PERFECT AND NON-PERFECT TEST SCORES

Test Scores	Program Ss	Conventional Ss	
Perfect	8	15	23
Non-Perfect	14	11	25
	22	26	48

Z= 1.184

Non-siginficant

APPENDIX K

SIGNIFICANCE TEST OF PROGRAM AND CONVENTIONAL Ss' PROBLEM I SOLUTIONS

Problem I	Program Ss	Conventional Ss	
Solved	18	19	37
Non-Solved	39	38	77
	57	57	144

$Z = 0$

Non-Significant

APPENDIX L

SIGNIFICANCE TEST OF PROGRAM AND CONVENTIONAL Ss' PROBLEM I AND PROBLEM II SOLUTIONS

Problems	Program Ss	Conventional Ss	
Solved I Only	18	19	37
Solved I & II	16	16	32
	34	35	69

Z= .13

Non-Significant

APPENDIX M

WILCOXON TEST ON PROGRAM AND CONVENTIONAL Ss' PROBLEM I SOLUTION TIMES

$$Z = \frac{199.5 - \frac{27\,(28)}{4}}{\sqrt{\frac{27\,(28)\,(55)}{24}}} = .252$$

Non-Significant

APPENDIX N

SIGNIFICANCE TEST ON ALL SOLVERS AND NON-SOLVERS OF PROBLEM I AND I.Q. GROUP

Problem I	Above 100 IQ	Below 100 IQ	
Solvers	30	7	37
Non-Solvers	36	41	77
	66	48	144

Z= 3.27

Significant at .01 level.

APPENDIX O

SIGNIFICANCE TEST ON ALL SOLVERS AND NON-SOLVERS OF PROBLEMS I AND II

Problem II	Problem I Solvers	Problem I Non-Solvers	
Solvers	32	6	38
Non-Solvers	5	71	76
	37	77	114

Z= 8.13

Highly Significant

DIRECTIONS:

Please learn the information in this booklet. It is necessary to learn it well and understand it thoroughly in order to solve the problem. To help you do this, some pages are labelled TEXT and tell you what you need to know to solve the problem. Each text page is followed by one or more question pages (QUESTION PAGE). Please write the answers to these questions before turning to the next page where the correct answers can be found (ANSWER PAGE).

If after seeing the correct answer you would like to check the text, it is all right to do so. Take each page as it comes and do not look ahead.

(Pages 25 and 27 of this text are in the lower section, last 2 pages.)

TEXT

Joe Doodlebug is a strange sort of imaginary bug who lives in a world where he can do certain things and cannot do others.

QUESTION PAGE

Joe is a gi______________ bug who cannot do everything he wants to do.

ANSWER PAGE

Joe Doodlebug is a strange bug who cannot do everything he wants to do.

QUESTION PAGE

Joe does some things but ________________ do everything he wants to do.

ANSWER PAGE

Joe does some things but cannot do everything he wants to do.

TEXT

He can jump in only four different directions: North, South, East, or West. He cannot jump diagonally: e.g., southwest, northeast, etc.

QUESTION PAGE

Joe can jump in only ______________ different directions.

ANSWER PAGE

Joe can jump in only four (4) different directions.

QUESTION PAGE

The directions are North, ______________, ______________, and

______________.

ANSWER PAGE

The directions are North, South, East, and West.

QUESTION PAGE

____________ other directions are excluded, including

di ______________________ jumps.

ANSWER PAGE

All other directions are excluded, including diagonal jumps.

TEXT

Once Joe starts in any direction, that is, North, South, East, or West, he must jump four times in that same direction before he can switch to another direction.

QUESTION PAGE

Joe is certainly a ____________________ sort of bug. Once he begins jumping one way he must jump _________ ____________ in that direction before changing directions.

ANSWER PAGE

Joe is certainly a strange (peculiar) sort of bug. Once he begins jumping one way he must jump four times in that direction before changing directions.

TEXT

Joe can only jump; he cannot crawl, fly or walk.

QUESTION PAGE

Jr. can slide, sidle, inch, roll, creep; that is, locomote in any different ways. TRUE OR FALSE? ____________

ANSWER PAGE

False

QUESTION PAGE

Joe can only ________________.

ANSWER PAGE

Joe can only jump.

TEXT

Joe can jump large distances or very small distances. Joe cannot turn around.

QUESTION PAGE

Joe can only jump very small distances. TRUE OR FALSE? ________

ANSWER PAGE

FALSE

QUESTION PAGE

He can jump ___________ distances as well as __________ ____________ distances.

ANSWER PAGE

He can jump large distances as well as very short distances.

QUESTION PAGE

Joe can rotate or twist in flight in order to turn around.

TRUE OR FALSE? ______________

ANSWER PAGE

FALSE

QUESTION PAGE

It is impossible for Joe to ________ ____________________.

ANSWER PAGE

It is impossible for Joe to turn around.

QUESTION PAGE

Joe is an im________________ bug who can only __________ and then only in ________ different ________________. Once he starts he must continue to jump f______ ___________ in this direction before changing ________________. Joe cannot __________ ________________.

ANSWER PAGE

Joe is an imaginary bug who can only jump and then only in four different directions. Once he starts he must continue to jump four times in this direction before changing directions. Joe cannot turn around.

TEXT

Joe has been jumping all over the place getting some exercise when his master places a pile of food three feet directly West of him.

QUESTION PAGE

Joe has just been doing a lot of jumping. TRUE OR FALSE?

____

ANSWER PAGE

TRUE

QUESTION PAGE

His master has just placed a pile of food __________ feet directly _________ of him.

ANSWER PAGE

His master has just placed a pile of food three feet directly West of him.

TEXT

Joe notices that the pile of food is a little larger than he. As soon as he sees the food he stops dead in his tracks facing North.

QUESTION PAGE

When Joe sees the food he is facing _____________.

ANSWER PAGE

When Joe sees the food he is facing North.

QUESTION PAGE

He realizes that the pile of food, just to the W________ of him, is l__________ than he is.

ANSWER PAGE

He realises that the pile of food, just to the West of him, is larger than he is.

TEXT

After all his exercise Joe is very hungry and wants to get to the food as quickly as he possibly can. Joe examines the situation and feels frustrated when he realizes that he will absolutely have to make four jumps and no fewer before he can get the food.

QUESTION PAGE

Joe is impatient to get to the food. TRUE OR FALSE? ____________

ANSWER PAGE

TRUE

45.

QUESTION PAGE

Joe has been exercising and is very ______________.

ANSWER PAGE

Joe has been exercising and is very hungry.

QUESTION PAGE

Since he is so hungry, Joe is ____________________ because he *has* to take four jumps to get the food.

ANSWER PAGE

Since he is so hungry, Joe is frustrated because he has to take four jumps to get the food.

QUESTION PAGE

Joe is frustrated because he has to take __________ jumps to get the food.

ANSWER PAGE

Joe is frustrated because he has to take four jumps to get the food.

QUESTION PAGE

Joe could not possibly reach the food in fewer than four jumps.

TRUE OR FALSE? ____________________

ANSWER PAGE

TRUE